On the Seasonal Forecasting of Regional Tropical Cyclone Activity

G.A. Vecchi[1,2], T. Delworth[1,2], R. Gudgel[1], S. Kapnick[2], A. Rosati[1], A.T. Wittenberg[1], F. Zeng[1], W. Anderson[1], V. Balaji[2], K. Dixon[1], L. Jia[1,3], H.-S. Kim[4], L. Krishnamurthy[1,3], R. Msadek[1,3], W.F. Stern[1], S.D. Underwood[5], G. Villarini[6], X. Yang[1,3], S. Zhang[1]

1. National Oceanic and Atmospheric Administration/Geophysical Fluid Dynamics Laboratory, Princeton, NJ, USA.
2. Atmospheric and Oceanic Sciences Program, Princeton University, Princeton, NJ, USA.
3. University Corporation for Atmospheric Research, Boulder, CO, USA.
4. Korea Maritime University, Korea.
5. D.R.C., NOAA/GFDL, Princeton, NJ, USA.
6. IIHR-Hydroscience and Engineering, The University of Iowa, Iowa City, Iowa, USA.

Submitted to *Journal of Climate*

18-February-2014

MANUSCRIPT IN PREPARATION

DO NOT QUOTE OR CITE WITHOUT PERMISSION

Abstract:

Tropical cyclones (TCs) are a hazard to life and property and a prominent element of the global climate system, therefore understanding and predicting TC location, intensity and frequency is of both societal and scientific significance. Methodologies exist to predict basin-wide, seasonally-aggregated TC activity months, seasons and even years in advance. We here show that a newly developed high-resolution global climate model can produce skillful forecasts of seasonal TC activity at spatial scales finer than basin-wide from months and seasons in advance of the TC season. The climate model used here is targeted at predicting regional climate and the statistics of weather extremes on seasonal to decadal timescales, and is comprised of a high-resolution (50km×50km) atmosphere and land components, and a more moderate (~100km) sea ice and ocean components. The simulation of TC climatology and interannual variations in this climate model is substantially improved by correcting systematic ocean biases through "flux-adjustment." We perform a suite of 12-month duration retrospective forecasts over the 1981-2012 period, after initializing the climate model to observationally-constrained conditions at the start of each forecast period – using both the standard and flux-adjusted versions of the model. The standard and flux-adjusted forecasts exhibit equivalent skill at predicting NH TC season sea surface temperature, but the flux-adjusted model exhibits substantially improved basin-wide and regional TC activity forecasts, highlighting the role of systematic biases in limiting the quality of TC forecasts. These results suggest that dynamical forecasts of seasonally-aggregated regional TC activity months in advance are feasible.

1. Introduction:

Predicting and projecting future tropical cyclone (TC) activity is a topic of scientific interest and high societal significance. Forecasts of TCs provide information to support planning, with the potential utility of the forecasts limited in part by their expected and realized skill and by the relevance of the quantity being predicted to the particular decision structure. A variety of methodologies have been developed to predict the path and intensity of individual TCs days in advance and, because of their demonstrated skill and regionally-specific information, a broad range of sectors regularly implement decisions based on these one-to-five day forecasts. Given the potential utility of TC predictions on longer lead times, various methodologies have been developed to skillfully predict seasonally aggregated, basin-averaged indices of TC activity (*e.g.*, Vitart and Stockdale 2001; Vitart 2006, Vitart *et al.* 2007l Camargo et al. 2007.b, Smith et al. 2010, LaRow et al. 2010, Klotzbach and Gray 2009, Jagger and Elsner 2010, Alessandri et al. 2011, Vecchi et al. 2011, 2013, Villarini and Vecchi 2013). TCs have a range of impacts, which vary regionally (*e.g.,* Pielke Jr. et al. 2008; Kam et al. 2013; Villarini et al. 2014.a-.b; Scocimarro et al. 2014), and basin-wide TC activity can often be a poor indicator of activity in sub-regions of the basin – including coastal areas (*e.g.,* Villarini et al. 2011, 2012; Vecchi and Villarini 2014). The utility of seasonal TC forecasts to decision support would therefore be enhanced if seasonal TC activity on scales finer than basin-wide could be skillfully predicted. In addition, seasonal forecasts of regional TC activity would provide tests of the hypothesized controls on regional TC activity, and enable refinement of our understanding of and ability to project multi-decadal changes in regional TC activity (*e.g.,* Murakami and Wang 2010; Murakami *et al.* 2011, 2012, 2013, 2014;

Knutson *et al.* 2008; Bender *et al.* 2010). High-resolution dynamical models provide a potential framework in this direction if they can represent and predict large-scale climate conditions and the processes that connect them to regional TC activity.

In general, one can view the TC forecast problem as a two-step process: 1) predict what the state of the future climate system is liable to be (the climate forecast), and 2) predict what the response of basin-wide TC frequency to what the future climate state is liable to be (the TC forecast). Sometimes the two steps occur within a single process, explicitly as when dynamical coupled climate models are used to predict the state of future climate, and the response of the TC-like vortices in the models is used to estimate future TC activity (*e.g.,* Vitart 2006; Smith et al. 2010), or implicitly when a statistical relationship between conditions prior to the TC season and the future season's TC activity is used (e.g., Gray 1984, Elsner and Jagger 2006, Klotzbach and Gray 2009). Since both the evolution of the climate system and the response of TC activity to climate are chaotic processes, these forecasts are not generally deterministic (*i.e.*, giving a single number) but probabilistic (*i.e.,* describing the probability of a range of plausible outcomes). Methodologies using a two-step approach to forecasting basin-wide activity include high-resolution dynamical model forecasts forced with either predicted or persisted climate anomalies (e.g., Zhao et al. 2009; LaRow et al. 2010, 2013; Chen and Lin 2011, 2013) and hybrid statistical-dynamical methods for seasonal TC forecast (e.g., Wang *et al.* 2010; Vecchi et al. 2011, 2013.a, 2014; Villarini and Vecchi 2013). These various methodologies have advantages and disadvantages relative to one another, but have all been shown to be potentially skillful at predicting basin-wide activity.

Large-scale climate variations and changes impact seasonal TC activity by impacting the environment in which TCs form, develop, propagate, and dissipate (*e.g.,* Gray 1984; Emanuel 1995; Bister and Emanuel 1998; Emanuel and Nolan 2004; Camargo et al. 2007.c, 2014; Knutson et al. 2010, 2013; Zhao et al. 2009; Vecchi and Soden 2007; Kossin and Vimont 2007; Vimont and Kossin 2007; Emanuel *et al.* 2008; Vecchi et al. 2008; Bender *et al.* 2010; Villarini et al. 2010, 2011, 2012; Tippett et al. 2012; Msadek et al. 2014.b). Climate models of moderate and high resolution can simulate aspects of both large-scale climate variations relevant to TCs (e.g., Broccoli and Manabe 1990; Vitart *et al.* 1997; Emanuel et al. 2008; Knutson et al. 2008, 2013; Vecchi and Soden 2007.a, 2007.b, et al. 2007.c, 2014; Wang et al. 2010; Vecchi et al. 2011), as well as aspects of the response of TCs to these climate changes (e.g., Knutson et al. 2008, 2013; LaRow et al., 2010; LaRow 2013; Zhao et al. 2009, 2010; Wang et al. 2014). However, climate models have deficiencies in both their large-scale climate as well as in the mean distribution of TCs. It has been hypothesized that large-scale model biases could be behind some of the model biases in TC simulation and sensitivity to climate (*e.g.*, LaRow *et al.* 2013; Kim *et al.* 2014; Murkami *et al.* 2014).

A range of observational and modeling studies indicate that aspects of the seasonally-aggregated TC activity at spatial scales finer than basin-wide are influenced by large-scale atmospheric and oceanic conditions (*e.g.,* Elsner *et al.* 2001; Camargo *et al.* 2007.a., 2008; Kossin *et al.* 2010; Murakami and Wang 2010; Villarini *et al.* 2010, 2012, 2014; Murakami *et al.* 2011, 2013; Colbert and Soden 2012; Zhang *et al.* 2012, 2013.a-.c;

Colbert *et al.* 2014; Kim *et al.* 2014) – including modes of climate variability that are potentially predictable months in advance, such as the El Niño-Southern Oscillation (ENSO) phenomenon and the Atlantic Meridional Mode (AMM), and the response of climate to radiative forcing changes. Therefore, we hypothesize that there is predictability to the regional structure of TC activity at scales finer than basin-wide. Further, we hypothesize that initialized predictions with a high-resolution coupled climate model are one way of extracting this predictable information. Finally, we hypothesize that biases in large-scale climate limit the simulation and forecast skill for TC activity, and that improvements to large-scale model biases will improve the simulation and prediction of TC activity in a high-resolution modeling system.

Here we use a recently developed high-resolution (~50km atmosphere and land resolution) coupled climate model to test the above hypotheses through climate simulations and initialized seasonal predictions. We assess the ability of the model to predict regional TC activity in the Northern Hemisphere (NH) Pacific and Atlantic Oceans, on multi-season leads. We also assess the impact of model biases that originate from biases in sea surface temperature (SST) on the simulation and seasonal forecast of TCs, by exploring parallel experiments with a free-running model and a version of the model whose fluxes are modified to bring its climatological SST in closer alignment with observations ("flux adjustment" using the methodology of Magnusson et al. (2013)).

In the next section we describe the models used, the forecast experiments, and ways of estimating and assessing TC activity. In Section 3, we present the results, focusing first

on the ability of different configurations of the free-running model to capture TC activity, then on the ability of the model to predict SST, and basin-wide and regional TC activity. In the final section we offer a summary of the results and some concluding remarks.

2. Methods:

a. Observational data

We use version v03r04 of the International Best Track Archive for Climate Stewardship (IBTRACS; Knapp et al. 2010) as our reference tropical cyclone dataset. To build consistency with the model-based definition of tropical cyclones, which has an explicit duration threshold (Section 2.f.ii), when comparing against model TC tracks with a two-day (or three-day, briefly in Section 3.a) duration threshold we consider only those storms for which winds exceed gale force and are classified as either topical or subtropical for over eight (twelve) 6-hourly best track fixes. We multiply the 1-minute maximum wind speeds archived in IBTRACS by 0.88 to estimate the 10-minute maximum wind speeds (Knapp et al. 2010).

We explore three main monthly SST datasets: the United Kingdom's Meteorological Office Hadley Centre's Interpolated SST product (HadISST.v1; Rayner et al. 2004), the European Centre for Medium Range Weather Forecasting's Interim Reanalysis SST dataset (Dee *et al.* 2011), and SST from the National Aeronautics and Space Administration (NASA) Modern-Era Retrospective Analysis for Research and Applications (MERRA) reanalysis (Reinecker et al. 2012). We use the HadISST.v1 SST, climatological sea surface salinity (SSS) from the World Ocean Atlas 2005 (Levitus et al.

2005) and surface zonal and meridional wind stresses from ERA-I to build our "flux adjusted" version of the model (Section 2.c). In addition, we use three-dimensional atmospheric temperature, wind and humidity data from the ERA-I and MERRA analyses as estimates to assess the large-scale structure of the atmosphere in the model simulations (Section 3.a).

b. Model description

In order to build a seasonal-to-decadal forecast system for regional climate impacts, including TCs, we have built a high-resolution coupled climate model, with its high resolution focused on the land and atmosphere components. The atmosphere and land components of this model are taken from the high-resolution Coupled Model version 2.5 (CM2.5; Delworth et al. 2012) recently developed at the Geophysical Fluid Dynamics Laboratory (GFDL), with a horizontal resolution of approximately 50km×50km using a cubed sphere finite volume dynamical core (Putnam and Lin 2007). However, in contrast to CM2.5, which has high resolution in both its atmosphere and ocean components, the ocean and sea ice components of this new model are based on the low resolution GFDL Coupled Model version 2.1 (CM2.1, Delworth et al. 2006; Wittenberg et al. 2006; Gnanadesikan et al. 2006). CM2.1, which has a horizontal grid spacing of 1° for the ocean and sea ice components (telescoping to 0.333° meridional spacing near the equator), and ~2° for the atmosphere and land components, has been used for numerous seasonal-to-decadal variability research, predictability and forecast activities (Vecchi *et al.* 2006, 2011, 2013; Zhang *et al.* 2007; Song *et al.* 2008; Wittenberg 2009; Msadek *et*

al. 2010, 2013, 2014.a; Choi *et al.* 2013; Yang *et al.* 2013; Kosaka *et al.* 2013; Wittenberg *et al.* 2014.a).

The new coupled climate model used here is referred to as the Forecast-oriented Low Ocean Resolution version of CM2.5, or FLOR. Our goal of capturing regional scales and extreme events (including TCs) requires us to pursue a model with high atmosphere and land resolution. The relatively lower ocean/sea ice resolution provides computational efficiency relative to the full version of CM2.5, allowing us to pursue large ensembles of forecasts. A coupled Ensemble Kalman Filter (EnKF) data assimilation system was built on CM2.1 (Zhang et al. 2007), which underpins our quasi-operational intraseasonal to decadal forecast activities. So an additional benefit of using the low ocean resolution in FLOR is that we can readily take ocean and sea ice initial conditions from the CM2.1 EnKF, which are key sources of predictability on multi-month to multi-season leads. A coupled assimilation project with FLOR is underway, which we expect will yield further improvements over the performance reported here.

The high-resolution CM2.5 model, which includes enhanced resolution in both its atmosphere/land and ocean/sea ice components, exhibited substantial improvements in its near-surface and atmospheric climate simulation relative to CM2.1 (*e.g.,* Delworth *et al.* 2012, Doi *et al.* 2012; Delworth and Zeng 2014; Wittenberg et al. 2014.b). In building FLOR, we hypothesize that the improvements in the simulation by CM2.5 of the climate features that are crucial to the forecast of seasonal-to-decadal regional climate and extremes arise from enhancements to atmosphere/land rather than ocean/sea ice

resolution. For the simulation of a series of near-surface and atmospheric quantities, such as the structure of anomalies tied to the ENSO phenomenon, and large-scale SST, land and ocean precipitation and near-surface winds, the improvements seen in CM2.5 relative to CM2.1 are evident in FLOR (Jia *et al.* 2014; Wittenberg *et al.* 2014.b). This suggests that, at least for the range of horizontal resolutions we have explored (between 1° and 0.1° for the ocean/sea ice, and 250km and 50km for the atmosphere/land), and for the numerical methods and parameter settings in these models, improvements in the simulation of near-surface climate and its variability are more closely connected to atmospheric than oceanic resolutions. This is a fortuitous result, since the cost of running FLOR is about half of that for the full-blown CM2.5 and we already have ocean/sea ice initial conditions at the resolution of FLOR.

We have explored two alternative versions of FLOR, which are referred to internally at NOAA/GFDL as FLOR-B01 and FLOR-A06. Since most of the results described in this paper, along with the flux-adjusted version of the model, are done using FLOR-B01, henceforth we will refer to that version of the model as "FLOR", without the modifier. The alternative formulation of FLOR will be referred to as FLOR-A06. These two model versions have identical atmospheric, land and sea ice configurations, but have slightly different parameterizations in the ocean. In both versions of FLOR, the ocean component has been slightly altered from that of Delworth *et al.* (2006) version of CM2.1 by having a more realistic representation of the solar absorption by the ocean, using a biharmonic horizontal viscosity scheme, as well as some fixes documented in Delworth *et al.* (2012). In addition to these changes, FLOR-B01 incorporates the newer, higher order advection

scheme used by CM2.5 (Delworth et al. 2012) and an updated parameterization for eddies (Ferrari et al. 2012).

The resulting models, FLOR and FLOR-A06, have most of their computational expense and resolution concentrated in the atmosphere and land components.The choice to concentrate resolution in the atmosphere/land, and keep the ocean resolution relatively low, had three principal motivations: i) FLOR is being targeted to understanding and predicting regional climate and extremes, for which atmosphere and land resolution are likely to be of value; ii) computational constraints limited the ensemble sizes and length of experiments that could be performed with the full, high-ocean-resolution CM2.5; and iii) ocean and sea ice initial conditions over the period 1980-2013 are available on the resolution of CM2.1, making the generation of initialized experiments relatively straightforward. A further consideration was the quality of the simulation near-surface and atmospheric climate, which was found to improve considerably as the atmospheric resolution went from ~2° in CM2.1 to 0.5° in FLOR (Jia *et al.* 2014; Wittenberg *et al.* 2014.b), yielding approximately 20 atmospheric grid points for every previous grid point. However, various measures of improvement to near surface and atmospheric climate showed much more marginal improvements from the additional resolution in the ocean going from CM2.5 to FLOR (Jia *et al.* 2014; Wittenberg *et al.* 2014).

c. Flux adjustment

We wish to test the hypotheses that: (i) improvements to the mean climate simulation should lead to improvements in the simulation of TCs (*e.g.,* Kim et al. 2014), and (ii) that

an improved mean simulation of TC activity should yield improved forecasts of basin-wide and regional TC activity. In order to test these hypotheses, we developed an alternative configuration of FLOR whose resolution, numerics and parameter settings are identical to the standard FLOR configuration, except it is "Flux Adjusted." That is, climatological adjustments are made to the model's momentum, enthalpy and freshwater fluxes from atmosphere to ocean to bring the model's long-term climatology of SST and surface wind stress closer to observational estimates over 1979-2012. Flux adjustments are computed applying a method similar to that of Magnusson et al. (2013). We refer to this alternative configuration as FLOR-FA.

The procedure we follow to build FLOR-FA is the following, which begins from the end of a 100-year control simulation with FLOR using 1990-levels of radiative forcing and land use:

- A simulation with FLOR is performed over 1961-2012, restoring the model's SSS to the World Ocean Atlas climatological values (Levitus et al. 2005) and SST to the 1961-2012 monthly estimates from the UK Met Office Hadley Center SST product. The SSS and SST values are restored using a 5-day restoring timescale and this experiment is referred to as FLOR-NUDGE1.
- The output of FLOR-NUDGE1 is compared to the ERA-I over 1979-2012 to compute monthly climatological differences in the zonal and meridional momentum flux between atmosphere and ocean. These climatological differences will be referred to as TAU_ADJUST.

- The nudging experiment is repeated, this time adding the climatological TAU_ADJUST to the FLOR simulation while SSS and SST are restored to the observational estimates. This experiment is referred to as FLOR-NUDGE2.
- The climatological SSS and SST adjustments over 1979-2012 are computed from FLOR-NUDGE2, with their global-mean, annual-mean removed. These adjustments are referred to as SSS_ADJUST and SST_ADJUST.
- The final flux-adjusted experiment is performed by adding the climatological TAU_ADJUST, SSS_ADJUST and SST_ADJUST to FLOR. This produces the final simulation referred to as FLOR-FA.

In addition to our standard FLOR-FA model derived as described above, we tested an intermediate version in which we only adjusted enthalpy and freshwater fluxes, after nudging the observational estimates. This alternative flux adjusted model, which we will refer to as FLOR-FA.05, exhibits comparable performance to the standard FLOR-FA and is used briefly in Section 3.d to assess impacts of ensemble size on prediction skill. Both FLOR-FA versions are based on FLOR-B01.

d. Control simulations

We generate 100-year control climate simulations with both configurations of the FLOR model (standard and flux-adjusted) by prescribing radiative forcing and land-use conditions representative of 1990. These experiments are referred to as "Present Day Control" experiments with FLOR and FLOR-FA. These experiments are used to

characterize the climatological simulation and interannual variability of FLOR and FLOR-FA.

e. Forecast experiments

We explore this question through a series of 12-member ensemble retrospective seasonal forecasts initialized 1-January, 1-April and 1-July over 1981-2013, each integrated for 12 months with each version of the model – or 2,376 model-years of retrospective forecasts. Because FLOR has an ocean and sea ice component on the same grid as CM2.1, which is our current "workhorse" seasonal-to-decadal forecast model at GFDL and for which we have a set of initial conditions built through EnKF data assimilation, for each forecast we initialize each of the 12 ensemble members with an ensemble member of the CM2.1 EnFK ocean and sea ice initial conditions. For our atmosphere and land initial conditions we use initial states from a suite of SST-forced atmosphere-land only simulations using the components in FLOR. That is, the ocean and sea ice are initialized with observationally-constrained estimates of their state, while observations impact the atmosphere and land initial state only through the information that is contained in the SST and radiative forcing that is used in the atmospheric general circulation model (AGCM) experiments. Since proper initialization is a key source of seasonal predictability, the experiments described here are not "optimal" forecast experiments, but represent a lower bound, to some extent, on the potential retrospective predictive skill of a system like FLOR. However, retrospective forecasts often outperform real forecasts – even when care is taken to cross-validate – so these experiments are not necessarily a lower bound estimate on future forecast skill. We pursue this suboptimal experimental

design since it allows us to efficiently assess aspects of the performance of FLOR and provides a baseline for future experiments using an assimilation system built with FLOR. Further, since the initial conditions are the same between our seasonal to decadal forecast system built with CM2.1 and these FLOR experiments, we can isolate the impact of model configuration on forecast skill.

Ensemble forecasts over the period 1981-2013, initialized on the first day of every month, are generated with both FLOR and FLOR-FA by using the ocean and atmosphere initial conditions generated from a Coupled EnKF analysis with CM2.1 (Zhang *et al.* 2007), which blends ocean and atmosphere observations into a coupled simulation. There is an ensemble of 12 ocean and sea ice initial conditions available over the period 1981-2013, each representing an equally-plausible state that is consistent with both the observed record and the climate model. Since the FLOR atmosphere and land models are different from those of CM2.1, we generate a series of atmosphere and land initial conditions offline by performing an ensemble of three SST-forced free-running AGCM simulations with the atmosphere/land component of FLOR. For the FLOR and FLOR-FA forecasts a 12-member ensemble is generated by applying the first AGCM member to the first four ocean members, the second AGCM member to ocean members five through eight, the third AGCM member to ocean members nine through twelve. We note that this initialization does not constrain the atmosphere beyond the information present in SST and radiative forcing, and that the ocean initial conditions are not "optimal" for FLOR or FLOR-FA. Therefore, we speculate that subsequent forecasts based on initial conditions

from an EnKF assimilation with FLOR and FLOR-FA, and that included atmospheric observations, are likely to improve on the solutions presented here.

Prediction experiments were also performed every month with FLOR-A06, and initialized on 1-July with an alternative FA version of FLOR in which only freshwater and enthalpy fluxes were corrected (referred to as FLOR-FA.05; Section 2.c); these two additional sets of predictions are only discussed briefly in Section 3.d as a way to assess the impact of increased ensemble size on forecast performance.

f. Tropical cyclone statistics:

i) Tropical cyclone tracking

Based on six-hourly snapshots of atmospheric state, we use the method described in Zhao et al. (2009), with the parameter settings in Kim *et al.* (2014), to track tropical cyclones in the FLOR output. This tracking scheme derives from the Vitart *et al.* (1997) tracking scheme. For most of our analyses we impose a two-day duration threshold on TCs before they are identified, and thus compare to observations with a similar duration threshold applied – since the history of counts of TSs of duration shorter than 2 days does not correspond to that of longer duration storms (Landsea *et al.* 2010; Villarini et al. 2011.a). To define TCs of different categories (e.g., tropical storms, Category 1 cyclone, etc.), we use a 90% scale on the observed threshold to account for the model resolution, based on Walsh *et al.* (2010) – so the model threshold for gale force winds is 15.3 $m \cdot s^{-1}$, rather than 17 $m \cdot s^{-1}$. When exploring basin-wide counts in the retrospective forecasts, model

counts are scaled by the ratio of the observed to ensemble-mean predicted values for the period 1982-2005:

$$C^*(t,e) = \frac{\langle O \rangle_{1982-2005}}{\langle \bar{C} \rangle_{1982-2005}} C(t,e)$$

where *C(t,e)* is the raw count prediction for year *t* and ensemble member *e*, $\langle . \rangle_{1982\text{-}2005}$ is the time-average over 1982-2005, and the overbar denotes ensemble averaging.

ii) Tropical cyclone density

We use "TC density" as a metric with which to assess the predictability of regional TC activity; we define TC density as the total number of days in a season in which a TC is inside a box 10° longitude by 10° latitude, centered in each 1° gridpoint. We explore 10°×10° regions because they are smaller than the scale of the basins, but still large enough to have a sufficiently large sample size to perform meaningful statistics with 32 years of verification data. We compute 10°×10° density at every point in a 1°×1° grid to minimize the impact of the edges of larger discrete boxes in computing density (*e.g.*, a storm passing at a position just slightly to the east of an edge and one just slightly to the west of an edge would be placed in two disjoint 10° boxes; having sliding boxes reduces this impact). The 10° scale is comparable to the average diameter of observed TCs (measured by the outer radius of the TC; Chavas and Emanuel, 2010), and is broad enough to include most of the areas where impacts of individual TCs in models and observations are evident (*e.g.,* Lin *et al.* 2010; Villarini and Smith; Villarini *et al.* 2011.b, 2014.a-.b; Scocimarro *et al.* 2014).

iii) Statistical-dynamical hybrid scheme:

The main focus of this work is the seasonal forecasting of regional NH TC activity, but in order to assess the performance of this new system against its predecessor system (CM2.1), we explore predictions of North Atlantic hurricane frequency. We use the hybrid statistical-dynamical North Atlantic hurricane frequency forecast framework by Vecchi *et al.* (2011, 2013), referred to as HyHUFS (Hybrid Hurricane Forecasting System), to compare the North Atlantic basin-wide hurricane forecasts of FLOR and FLOR-FA, to the forecasts using CM2.1. The HyHUFS scheme combines a statistical emulator of a high-resolution dynamical atmospheric model (Zhao *et al.* 2009, 2010) and initialized forecasts of SST. The statistical emulator is formulated as a Poisson regression model with two predictors: Tropical Atlantic SST and Tropical-mean SST, each averaged over the August-October season. The choice of these two predictors is motivated by dynamical considerations, observed relationships between hurricane activity and SST, and the sensitivity of dynamical models to SST perturbations (*e.g.,* Vecchi and Soden 2007; Swanson 2008; Vecchi *et al.* 2008, 2013.b; Knutson *et al.* 2008, 2013; Villarini *et al.* 2010, 2012; Vecchi and Knutson 2011; Tippett et al. 2011; Camargo et al. 2012). Following Vecchi *et al.* (2011, 2013.a), we model the rate of occurrence λ of North Atlantic hurricane frequency using a Poisson regression model as follows:

$$\lambda = e^{1.707+1.388 SST_{MDR} -1.521 SST_{TROP}} \qquad \text{(Eq.1)}$$

where SST_{MDR} and SST_{TROP} are anomalies in the regional SST indices relative to the 1982–2005 average. SST_{MDR} is the average over the hurricane main development region (80°W-20°W, 10°N-25°N), and SST_{TROP} is the global, 30°S-30°N average of SST.

3. **Results**:

a. Simulation of TC Activity

Although the focus of this manuscript is TC activity forecasts in the NH Pacific and Atlantic, we begin by briefly exploring the global geographic distribution TCs in FLOR. The Present Day Control simulation with FLOR is able to recover many aspects of the geographic distribution of genesis and storm track (Fig. 1.b-e) that bears considerable resemblance to the observed (Fig. 1.a.-.d), yet biases in the simulation of TCs in FLOR are evident. For example, there is too much activity in the Southern Hemisphere and Indian Ocean. There are also regional biases in the NH Pacific and Atlantic basins, which are the main focus of this work. In the Northern Central Pacific (around the Hawaiian Islands) there is excessive activity in FLOR – such that the clear distinction between the East and West Pacific in observations is not evident in FLOR. In the North Atlantic there is no genesis in the Caribbean and Gulf of Mexico, and very few tracks make it into the Western Atlantic.

Overall, the simulation of TCs in FLOR is comparable to that in CM2.5 (Kim *et al.* 2014), although there is more North Atlantic activity in FLOR than in CM2.5. Based on a 3-day duration threshold Kim *et al.* (2014) report 2.4 TCs/year in the North Atlantic, using the same duration threshold FLOR has 4.5 TCs/year (the observed average is 7.3 over the 1981-2011 period and 6.7 over the 1966-2011 period). The annual cycle of genesis in each basin is of comparable quality to that of CM2.5, comparing well with observations in all basins except the North Indian Ocean (not shown). A key deficiency both in FLOR and CM2.5 is in the intensity distribution of the TCs: in part due to their

resolution, both models have too small a range for TC intensity, which is truncated away from high intensity.

It has been hypothesized (Kim et al. 2013, LaRow 2013) that biases in model simulations of TCs arise in part due to large-scale climate, some of which may be traced to biases in the SST simulation. The standard version of FLOR exhibits substantial SST biases during NH TC season (July-November, Fig. 2.b), with cold biases in the North Atlantic and Northwest Pacific, and warm biases near the equator. As can be seen in the middle panels of Fig. 3, FLOR exhibits considerable vertical wind shear and potential intensity (PI; Bister and Emanuel 1998). High values of wind shear tend to limit TC development and intensification (e.g., Frank and Richie 2001; Emanuel and Nolan 2004), while high values of PI tend to enhance TC development (e.g., Bister and Emanuel 1998; Emanuel 2008). The PI and shear biases in FLOR would tend to make the North Atlantic, in particular the western sector of the Atlantic, anomalously hostile to TC genesis and intensification Further, FLOR exhibits a low-shear, high-PI region in the North Central Pacific, which would act to make that region overly favorable to TC genesis and intensification.

The Present Day Control simulation with the flux-adjusted configuration of FLOR allows us to test the hypothesis that improved representation of mean climate should lead to improved representation of TC climatology. As designed, the flux adjustments reduce the climatological biases in SS, leading to long-term average SST biases in NH TC season (July-November) SST of generally less than 0.5°C - when the standard FLOR model can have biases of many °C – even exceeding 3°C over large regions (Fig. 2.c). As a result of

the reduced SST biases, the FLOR-FA model has a substantially improved simulation of many aspects of near-surface climate, including vertical wind shear and PI over the NH TC season (lower panels in Fig. 3).

Concurrent with improvements in NH PI and wind shear, the climatology of Pacific and Atlantic TC genesis and tracks in FLOR-FA is improved relative to the standard version of FLOR (Fig. 1). The flux adjustments cause the western North Atlantic to be less hostile to TCs, allowing TC genesis and track to extend into the Caribbean, Gulf of Mexico and Sargasso; in FLOR-FA there is a clear – and more realistic – separation between Eastern and Western North Pacific TCs. These results lend support to the hypothesis that NH Pacific and Atlantic TC frequency and track simulation depends, to a substantial degree, on improved simulation of large-scale climate.

However, the FA run does not improve all aspects of TC simulation in FLOR. In particular, FLOR-FA does not produce substantial improvement in the simulation of Indian Ocean and Southern Hemisphere TC climatology (Fig. 1) – suggesting that factors beyond those addressed through flux adjustment are important to correctly simulating TCs in these regions. Currently, we are exploring a set of possibilities for misrepresented processes in the model that could be behind these persistent TC biases. We suspect that the spuriously enhanced convection in the southern hemisphere tropics (known as the "Double Inter-Tropical Convergence Zone" error) that has been pervasive in dynamical models for decades is partly to blame for these TC errors, the underlying causes for which remain elusive.

We now focus more closely on the simulation of tropical cyclone density in the NH Pacific and Atlantic by comparing 10°×10° TC density in observations (Fig. 4.a) and the FLOR models (Fig. 4.b.-.c). Consistent with the TC track maps in Fig. 1, FLOR (Fig. 4.b) has excessive activity in the north Pacific, particularly in the central and western sections, and almost no activity in the Western Atlantic. The flux-adjusted version of FLOR (Fig. 4.c) shows considerable improvement over FLOR in the Atlantic, and some improved representation of the separation between East and West Pacific TC basins. However, FLOR-FA still has too much activity in the Pacific relative to the Atlantic – a deficiency seen in other models at GFDL, even when forced with observed SSTs (*e.g.,* Zhao *et al.* 2009, 2010; Chen and Lin 2012).

Figure 5 shows the rank correlation of TC density to NIÑO3.4 SST anomalies (SSTA) in observations and the Present Day Control Simulations of FLOR and FLOR-FA. The observed record indicates that TC density in the West Pacific shows a strong positive relationship to El Niño, with weaker positive correlations in the East Pacific and negative ones in the Atlantic (Fig. 5.a). To some degree FLOR recovers some of the basic features seen in observation, with positive correlations in the West Pacific and negative correlations in the Atlantic (Fig. 5.b). However, FLOR also exhibits differences with observations: the region of positive correlation in the West Pacific is displaced about 20°-40° to the east relative to observations, the negative correlation values in the North Atlantic are larger than observed (and there is insufficient activity to compute a correlation in the western Atlantic), and the far East Pacific shows large negative

correlations that are absent in observations. Meanwhile, the correlations of TC density to NIÑO3.4 in FLOR-FA agree more with observations than do those in FLOR (Fig. 5.c). Flux adjustment appears to improve the sensitivity of TC activity to climate variability, in addition to improving aspects of the mean TC climatology.

We speculate that the differences in relationship of TC activity to El Niño in FLOR and FLOR-FA may be in part due to the differences in the character of El Niño in each version of the model. The amplitude of El Niño in FLOR is substantially larger than that in observations and in FLOR-FA (Fig. 6; Wittenberg et al. 2014), including a larger number of "extreme" El Niño events in which atmospheric convection makes its way across to the eastern equatorial Pacific (*e.g.*, Vecchi and Harrison 2006, Vecchi 2006, Lengaigne and Vecchi 2006). We hypothesize that the stronger El Niños in FLOR, with a more eastward extension to their convective anomalies, would lead to an enhanced negative response in the East Pacific and North Atlantic and an eastward extension of the West Pacific positive correlation. This hypothesis is currently being tested with a suite of perturbation experiments (Lakshmi Krishnamurthy, personal communication).

b. Forecast of August-October SST

We begin our analysis of the retrospective forecasts using FLOR and FLOR-FA by focusing on the retrospective skill for August-October SST (ASO-SST) over the period 1981-2012 (Fig. 7) – since August-October is the peak of TC activity in the NH. For the July-, April- and January-initialized forecasts highlighted in Fig. 7, FLOR and FLOR-FA exhibit comparable correlation forecasting ASO-SST, and both exhibit skill that is either

comparable to or somewhat better than CM2.1. For all three models, forecast skill for ASO-SST is larger for shorter leads (forecasts initialized 1-July and verifying 1-August through 31-October) than for longer lead forecasts (initialized 1-April and 1-January) – as one would expect. Improvements relative to CM2.1 are most prominent in the western equatorial Pacific, at the edge of the observed West Pacific warm pool – a key location for the generation of the remote connections to tropical Pacific variations. As noted in Jia *et al.* (2014), FLOR and FLOR-FA show some improvement over CM2.1 in forecasts of eastern equatorial Pacific ASO-SSTs when initialized before boreal spring. Tropical Atlantic ASO-SST skill is comparable in all three systems, with slightly larger nominal correlation values in the FLOR-FA forecasts.

For January and April start dates, forecasts of ASO-SST in West Pacific regions of TC genesis exhibit substantially improved correlation in FLOR compared to CM2.1, with more modest improvements in the East Pacific and North Atlantic. A concern prior to generating these forecasts was the potential inconsistency between the ocean initial conditions generated using CM2.1 and the FLOR models; but any impact of that inconsistency is not sufficient to reduce the overall ASO-SST forecast skill with FLOR below that of CM2.1. We are thus encouraged to explore the ability of FLOR and FLOR-FA to predict seasonal NH Pacific and Atlantic TC activity from January, April and July initial conditions.

c. Forecast of Basin-wide TC Activity

As a first step in assessing FLOR's TC forecast skill, we focus on retrospective forecasts of Atlantic hurricane frequency. While our ultimate goal is forecasts of regional TC activity, basin-wide Atlantic hurricane frequency provides a useful touchstone. A hybrid statistical-dynamical forecast system for hurricane frequency based on CM2.1 has been developed (Vecchi *et al.* 2011; Section 2.e.i), and this hybrid system (HyHuFS) is readily applicable to forecasts of SST from any model - including FLOR and FLOR-FA. We can then compare the performance of the HyHuFS scheme in FLOR and FLOR-FA to that in CM2.1 (their predecessor model), and these can be compared to forecasts based on counting TCs directly in FLOR and FLOR-FA. We assess the 1981-2012 retrospective performance of these basin-wide NA hurricane frequency forecasts (Fig. 8) through the Spearman Rank Correlation (R_{rank}) and Mean Square Skill Score (MSSS), which provide complementary information about the performance of the forecast systems (Goddard et al. 2013). We use rank correlation as our correlation metric, since we do not expect the ensemble-mean forecast of number of hurricanes and the number of hurricanes observed each year (which is an integer count) to follow a Gaussian distribution. Rank correlation describes the ability of the forecast system to identify the relative ordering of years (least to most active) in the observed record correctly, while MSSS also includes information about the conditional bias of the forecasts. Both R_{rank} and MSSS have a value of 1 for a perfect forecast, with negative values indicating substantial failures in performance.

For most forecast initialization times, HyHuFS applied to FLOR and FLOR-FA SST forecasts performs as well or better than when applied to CM2.1 SST forecasts. For July-initialized forecasts CM2.1 HyHuFS has smilar retrospective R_{rank} than HyHuFS from

FLOR and FLOR-FA, but both FLOR and FLOR-FA outperform CM2.1 in MSSS – reflecting a larger conditional bias in the short-lead hybrid forecasts with CM2.1. For all leads, the HyHuFS forecasts with FLOR-FA SSTs show the best overall performance. Since HyHuFS is based on the scaled temperature difference between Atlantic and global tropical SST, FLOR-FA is able to successfully predict the difference between tropical Atlantic and tropical-mean SST in a way that leads to skillful Atlantic basin-wide hurricane forecasts from one- to three-season leads.

Comparing the darker blue bars and red bars in Fig. 8, representing the hybrid and dynamical forecasts, respectively, it is clear that the hybrid statistical-dynamical forecasts of Atlantic hurricane frequency outperform the purely dynamical forecasts based on counting TCs in both FLOR and FLOR-FA – at least at longer leads. This result may appear counterintuitive, yet is reasonable given the large amplitude of variations in hurricane frequency in the Atlantic that are unconstrained by SST (Zhao *et al.* 2009, 2010; Villarini *et al.* 2010, 2012) – in an SST-forced AGCM of comparable resolution to this, the standard deviation of NA hurricane frequency across ensembles forced with identical SST is 1.7 hurricanes per year (Zhao *et al.* 2009, 2010). Uncertainties in forecasts of hurricane frequency include an element arising from uncertainties in forecasts of large-scale climate (the two SST indices in HyHUFS, and the totality of the climate signal impacting hurricanes in the dynamical forecasts). The HyHUFS system predicts the expected value of hurricane frequency for each of the 12 ensemble members; the dynamical forecasts, on the other hand, give a single sample of hurricane frequency for each of the 12 ensemble members, so the estimates of the expected value of hurricane

frequency in these forecasts include a component from inadequately estimating the expected value for each ensemble member from a single realization. We suspect that these results may be general to some degree, and, for quantities with a large unforced component, properly designed hybrid statistical-dynamical models may be expected to outperform, and give a fuller representation of the forecast probability density than purely dynamical models, for the narrow questions to which their statistical elements are targeted. Recent analysis of FLOR forecasts of temperature and precipitation over land indicates that statistical refinement, essentially a reduced-space reconstruction of the predictands, leads to improvement over the raw forecasts (Jia et al. 2014). Therefore, statistical and dynamical forecast methodologies should not be viewed as competing alternatives, but efforts should be built to integrate them to build off the strengths of each.

From comparing the dynamical forecasts of North Atlantic hurricane frequency in FLOR to those in FLOR-FA (compare light red and dark red bars in Fig. 8), it is clear that the flux-adjustment leads to enhanced forecasts – particularly at longer leads. This is in part explainable by improvements in forecasts of large-scale conditions (*e.g.,* SST) in FLOR-FA (compare the skill of the hybrid forecasts in FLOR to FLOR-FA in Fig. 8). But there is an element of the improvement in FLOR-FA that comes from improved representation of the TC genesis and track structure in FLOR-FA, and the response of TC density to climatic variations – so that TCs tend to form and intensify in the correct position relative to climatological and anomalous large-scale climate conditions that impact their seasonal frequency. For July start dates, there is less of an improvement in dynamical Atlantic hurricane frequency forecasts between FLOR-FA and FLOR, as the models have been

initialized to conditions close to observations and there has been insufficient time for FLOR to have substantial drift to its own – more biased – climatology.

The improvement in climatological TC tracks in the FLOR-FA forecasts relative to forecasts with FLOR can be seen in Fig. 10. For the July-initialized forecasts, the climatological TC density in FLOR and FLOR-FA both match observations relatively well; both models have been initialized with observational estimates and in the few months between initialization and the end of the TC season, there is limited drift to the large-scale climate. However, for as lead-times for the forecasts become longer (April- and January-initialized forecasts), the TC density from the initialized forecasts with FLOR exhibits clear indications of the drift towards that model's free-running climatology. For the January forecasts, even if the FLOR forecasts had succeeded in recovering perfect large-scale anomalous conditions relevant to Atlantic hurricane variability, the model's TCs would be imperfectly aligned with those climate anomalies (unless they were spatially homogeneous anomalies). We hypothesize that this improvement in forecast skill of TCs from FA should also be evident in other quantities that exhibit strong nonlinearities (*e.g.,* features with genesis, limited existence and termination; features impacted by threshold nonlinearities), such as rainfall in arid regions, snowfall and midlatitude storms.

Given the improvement of North Atlantic seasonal hurricane frequency forecasts with FLOR and, in particular, FLOR-FA over CM2.1 (Fig. 8), we wanted to assess how the forecasts with this new model system compared with those in the published literature

(*e.g.,* Vitart *et al.* 2007; Klotzbach and Gray 2009; Zhao *et al.* 2009; LaRow *et al.* 2010; Wang *et al.* 2010; Chen and Lin 2013). Each of these other published studies used a different verification period, and they each focused on a different combination of start dates, so we compare the performance of the dynamical and HyHuFS predictions with FLOR-FA over the verification period and start dates used by each of the other systems (Fig. 9). In Fig. 9, symbols above the diagonal indicate nominal improved performance of FLOR-FA relative to the other methods. Overall, the performance of FLOR-FA is comparable to most of the other methods, with some indication that it outperformed the other systems at longer leads – particularly for the HyHuFS predictions with FLOR-FA. That is, not only does FLOR-FA outperform our old system (CM2.1-HyHuFS; Vecchi *et al.* 2011, 2013), but its performance is competitive relative to other published studies. It appears that differences in verification period are a small factor in the differences between retrospective skills in these various methods, so differences in correlation likely reflect differences in the forecast methods: compare the vertical span of like symbols (*e.g.,* circles) in Fig. 9, which indicates the dependence on verification period, with the horizontal span of like symbols, which indicates the dependence on method. However, retrospective performance is an imperfect estimate of future prediction skill.

Of particular interest is comparing FLOR-FA to the studies of Zhao *et al.* (2009; light green) and Chen and Lin (2013; violet), which were made using atmospheric models that share some elements with FLOR (namely the cubed sphere dynamical core of Putman and Lin, 2007). The method used in Chen and Lin (2013) differs from that in Zhao *et al.* (2009) by: i) using a higher resolution atmosphere (~25km instead of ~50km), ii)

initializing the atmospheric state with observational estimates, and iii) focusing on a different verification period. The verification period alone is unlikely to explain Chen and Lin's (2013) outperformance of Zhao *et al.* (2009), since the July-initialized FLOR-FA retrospective forecast skill is comparable for all verification intervals. Therefore, it appears that some combination of the enhanced resolution and atmospheric initialization played a role in the skill difference between Zhao *et al.* (2009) and Chen and Lin (2013), adding motivation to ongoing efforts to build a fully-coupled initialization system with FLOR/FLOR-FA.

d. Forecast of Regional TC Activity

We are encouraged to explore the predictive skill of FLOR for regional TC activity from its forecast quality for North Atlantic basin-wide activity (Fig. 8), NH SST (Fig. 6) and from its overall simulation of TC genesis and track climatology (Fig.s 1, 4 & 5). Variations of TC activity at spatial scales smaller than basin-wide have been connected to large-scale modes of climate variability that are potentially predictable on seasonal timescales, such as ENSO, the Atlantic Multi-decadal Oscillation, the Pacific Decadal Oscillation and the Atlantic Meridional Mode. Therefore, we expect that the initialized forecasts FLOR may exhibit skill in forecasts of regional TC activity. We further hypothesize that, particularly for longer leads when the model biases are able to emerge more fully, forecasts of regional TC activity with the flux-adjusted version of FLOR should outperform those with the standard version of FLOR. We expect FLOR-FA to outperform FLOR in regional TC activity forecasts both because of its improved

forecasts of basin-wide activity (Fig. 8) and because it has an improved track climatology.

For much of the NH Pacific and Atlantic basins there is significant skill in forecasts of regional TC activity initialized 1-July over the period 1981-2011 using FLOR and FLOR-FA (upper panels Fig. 11), measuring the retrospective performance of forecasts of regional TC activity using R_{rank}. The largest correlations tend to be in marine regions and at the margins of the modeled and observed TC density. There are significant retrospective correlations over some land areas, indicating the potential for some skillful seasonal forecasts of regional TC activity over land – although most land areas do not show skill.

The longer multi-season lead forecasts initialized in 1-April and 1-January show a rapid decrease in retrospective skill in the FLOR forecasts (left hand column Fig. 10), with only spotty regions of significant skill in January forecasts. However, FLOR-FA retains significant skill over broad areas for longer, with the January-initialized forecasts of regional TC activity in FLOR-FA comparable to those initialized in April in FLOR. Flux adjustment leads to substantial improvement in FLOR's ability to predict regional TC activity, though the skill near land decays rapidly for both FLOR and FLOR-FA. The strongest correlations, apparent over the longest leads, are evident in the West Pacific, generally collocated with the region exhibiting a strong connection to ENSO, including the narrow strip extending over Taiwan and southeastern China (Fig. 5). This collocation suggests that skillful ENSO forecast are likely to be behind the skill in the West Pacific.

The North Atlantic (centered in the Caribbean Sea and Western Gulf of Mexico) and Central Pacific regions of persistent skill are not regions with as strong a connection to ENSO as the West Pacific (Fig. 5), suggesting that skillful forecasts of other climate phenomena. We hypothesize that predictions of the Atlantic Meridional Mode are important for the North Atlantic skill (Vimont and Kossin 2007, Kossin and Vimont 2007), and that distinguishing between extreme and normal El Niño events (*e.g.,* Vecchi and Harrison 2006; Vecchi 2006; Lengaigne and Vecchi 2009) may provide some of the skill in the East and Central Pacific. These hypotheses are currently being tested.

The improvement in regional TC activity forecasts by flux adjustment is further highlighted in Fig. 12, which shows the fraction of the "TC regions" in the NH Pacific and Atlantic that exhibit significant (Fig. 12.a) or substantial (Fig. 12.b) retrospective rank correlation in the forecasts of TC density. At short leads (June- and July-initialization), the fraction of TC regions exhibiting significant skill is comparable in FLOR and FLOR-FA, but for longer leads there is a rapid divergence with FLOR-FA showing considerably larger areas with significant correlation. The fraction of TC regions with significant (at $p<0.1$) correlation in FLOR-FA forecasts initialized January (three-season lead) is larger than for FLOR forecasts initialized April (two-season lead); January-initialized forecasts with FLOR-FA have almost twice the area with significant rank correlation than do those with FLOR (Fig. 12.a). The difference between FLOR-FA and FLOR performance is more striking if one focuses on the percentage of TC regions that exhibit retrospective rank correlation exceeding 0.5 (Fig. 12.b); for all start dates FLOR-FA shows more area with rank correlation exceeding 0.5. With FLOR, flux

adjustment adds about a season of lead to the forecast performance of regional TC activity as measured by these two metrics.

This initial suite of forecasts with FLOR and FLOR-FA were performed with 12 ensemble members, which is likely sufficient for forecasts of large-scale ocean indices like NIÑO3.4. However, it is unclear the extent to which 12 ensemble members are sufficient for quantities with a large internal variability component like regional TC activity. It is possible that the skill in seasonal, regional TC forecasts described above may be enhanced through a larger ensemble set. In order to provide a preliminary assessment of the impact of larger ensemble sizes on the retrospective forecast skill of regional TC activity, we make use of the July-initialized forecasts that are available from four versions of FLOR (FLOR, FLOR-A06, FLOR-FA and FLOR-FA.05) and the observation that the July-initialized skill for TC density in all four versions is comparable, to generate a pseudo-48 member ensemble (Fig. 13). This 48-member ensemble should be compared to the 12-member ensemble with FLOR and FLOR-FA (Fig. 11.a-b). It is worth noting that the retrospective forecast performance of FLOR-A06 and FLOR-FA.05 in the quantities shown in Figs. 7, 8, 10 and 11 is comparable to that of FLOR and FLOR-FA, although there are differences in the ocean simulation of the various models and the spatial structure of their ENSO. Increasing ensemble size leads to systematic improvements in the performance of seasonal forecasts of regional TC activity, as can be seen through the red, yellow and green dots in Fig. 12.

Although in this test it appears that the large-scale gains from additional ensemble members are somewhat small (compare Fig. 13 to Fig. 11.a and 11.b), at this stage we are unable to assess the extent to which these results for July-initialized forecasts will hold for other leads (as the forecast performance of FLOR and FLOR-A06 degrades with lead more rapidly than FLOR-FA), or for a larger ensemble with FLOR-FA. Further, some of the regions in which there are increases in retrospective correlation from additional ensemble members are near land (*e.g.,* the northern Gulf of Mexico, the far western West Pacific, the far-eastern East Pacific), which could be of practical importance. We hypothesize that the nominally non-monotonic evolution of skill with lead-time in these predictions (*e.g.*, Fig. 12.b, comparing the skill in FLOR-FA in the Gulf of Mexico in Fig. 11) is due in part to the small ensemble size, and that a larger ensemble may make the forecasts skill decay more monotonically with lead time. Therefore, while the 12-member ensemble size was sufficient here to show the potential for seasonal forecasts of regional TC activity, we recommend larger ensemble sizes if possible, with lagged ensembles (*e.g.,* Vecchi *et al.* 2011, 2013) offering a potential way to create slightly larger ensemble sizes.

4. Summary and Discussion:

These initial retrospective forecasts of regional, seasonal TC activity with this high-resolution coupled climate model show skill across much of the NH Pacific and Atlantic basins multiple months in advance. In certain regions the flux-adjusted version of this model leads to significant regional skill multiple seasons in advance (Fig. 4). At all seasons, the rank correlations for regional TC activity are comparable to those seen with

basin-wide activity forecasts with these models (Fig. 3). Improvements in simulation of mean climate and TCs through enhanced resolution and flux-adjustment can lead to skillful retrospective forecasts of regional climate extremes, suggesting that future forecasts of these quantities may also be skillful.

Both FLOR and FLOR-FA produce somewhat realistic TC simulations in the NH Pacific and Atlantic basins – though deficiencies remain in both models. Overall, the simulation of FLOR-FA is superior to that of FLOR, indicating that improvements in the mean climatological SST improve simulation of TCs, either directly by improving the climatological simulation of large-scale conditions that impact TCs or indirectly by impacting the character of interannual variability.

Although these initial results are encouraging, these forecasts may be improved through a number of avenues. In these forecast experiments we did not attempt to initialize the atmosphere beyond the information that can be recovered from prescribing SST. Given the role of atmospheric patterns not necessarily linked to SST in modifying TC tracks (such as the role of the North Atlantic Oscillation in steering Atlantic TCs – Elsner et al. 2001; Kossin *et al.* 2010, Colbert and Soden 2012, Villarini *et al.* 2012, 2014), we suspect that atmospheric initialization of these modes may provide some additional improvement to these results. We have also used ocean and sea ice initial states built from a different model system; we are currently testing the hypothesis that, by providing an initial state more consistent with the underlying model, initial conditions generated within FLOR should enhance its skill in predicting large-scale and regional climate, and

the seasonal statistics of weather extremes (such as TCs). Further, our current ensemble size is 12, which is likely adequate for forecasts of large-scale climate indices (like ENSO indices), but may be inadequate for quantities with a large stochastic component – such are regional climate and the statistics of weather extremes. We are testing the impact of a larger ensemble size in improving forecasts of regional TC activity. This study was performed with two versions of a single climate model, and studies indicate that multi-model approaches can outperform forecasts using a single model. As climate models at resolutions comparable to ours are being run in multiple centers around the world (*e.g.,* Bell *et al.* 2013), the ability of different models and multi-model ensembles to outperform the results shown here should be explored.

As we noted, a statistical-dynamical hybrid approach outperformed the dynamical model at forecasts of basin-wide Atlantic hurricane frequency. The extent to which hybrid statistical-dynamical forecasts can improve on the results shown here should be explored. In particular, since forecasts of TC activity – particularly regional TC activity – are inherently probabilistic, it is important to develop appropriate error models for these regional TC forecasts. We suspect that the inter-ensemble spread of the forecasts is likely to be an inadequate error model, and efforts to build more adequate ones are paramount – because the utility of forecasts such as these will be limited by the absence of a suitable and reliable estimate of their uncertainty. For example, the results of Camargo *et al.* (2007.a, 2008), Kossin *et al.* (2010), Villarini *et al.* (2010, 2012, 2014.a), Colbert and Soden (2012), and Wang *et al.* (2012, 2013.a-.c) suggest some basis by which hybrid

models of regional TC activity could be built to complement and augment the purely dynamical results presented here. Efforts are underway to assess these strategies.

The analyses of seasonal predictions of regional TC activity in this manuscript have focused on deterministic measures of accuracy using the ensemble mean of the forecast as "best estimate." As was argued above and elsewhere (*e.g.*, Vecchi and Villarini 2014), climate predictions should be explicitly probabilistic. This study has not explicitly developed a probabilistic element to regional TC predictions, and doing so remains a priority for extensions beyond the present analysis. Future work should concentrate on building error models for the predictions of regional TC activity, and probabilistic assessments of the forecast performance. Such activities will likely lead to insights into the mechanisms controlling regional TC activity, as well as into its predictability, and are likely to yield much more reliable predictions. Large ensembles (with more than the 12 members presently available) are likely to be very useful in this process, providing an additional motivation for larger ensembles in future predictions (beyond the improvement in deterministic performance).

The results presented here show that skillful dynamical forecasts of seasonal regional TC activity at sub-basin scales are feasible months and seasons in advance – including in regions over and near land. The potential for these forecasts should be developed and enhanced, and their performance improved. Enhancements to models and understanding, and increased computer capacity, should enable these future developments.

Acknowledgements

We are grateful to E. Shevliakova and C. Gaitán for helpful comments and suggestions. This work is supported in part by NOAA's Climate Program Office, the National Science Foundation under Grant No. AGS-1262099 (Gabriele Villarini and Gabriel A. Vecchi), and by the Willis Research Network (Hyeong-Seog Kim). We are grateful to F. Vitart, P. Klotzbach, T. LaRow and H. Wang for providing data of the retrospective predictions skill of their systems.

References:

Alessandri, A., A. Borrelli, S. Gualdi, E. Scoccimarro, and S. Masina, 2011: Tropical cyclone count forecasting using a dynamical seasonal prediction system: Sensitivity to improved ocean initialization. *J. Climate*, **24**, 2963–2982.

Bell, R., J. Strachan, P.L. Vidale, K. Hodges, M. Roberts, 2013: *J. Climate*, **26**, 7966 doi: http://dx.doi.org/10.1175/JCLI-D-12-00749.1.

Bender, M. A., T. R. Knutson, R. E. Tuleya, J. J. Sirutis, G. A. Vecchi, S. T. Garner, and I. M. Held, 2010: Model impact of anthropogenic warming on the frequency of intense Atlantic hurricanes. *Science*, **327**, 454–458.

Bister, M., and K. A. Emanuel 1998: Dissipative heating and hurricane intensity. *Meteor. Atmos. Phys.*, **65**, 233–240.

Broccoli, A.J., and S. Manabe, 1990: Can existing climate models be used to study anthropogenic changes in tropical cyclone climate? *Geophys. Res. Lett.*, **17**, 1917-1920.

Camargo, S. J., Robertson, A. W., Gaffney, S. J., Smyth, P & and Ghil, M., 2007.a: Cluster analysis of typhoon tracks. Part II: Large-scale circulation and ENSO. *J. Clim.*, **20**, 3654–3676.

Camargo, S.J., A.G. Barnston, P. Klotzbach, and C. W. Landsea, 2007.b: Seasonal tropical cyclone forecasts. WMO Bull., 56, 297–309.

Camargo, S. J., K. A. Emanuel, and A. H. Sobel, 2007.c: Use of genesis potential index to diagnose ENSO effects upon tropical cyclone genesis, *J. Clim.*.

Camargo, S. J., Robertson, A.W., Barnston, A. G. & Ghil, M., 2008: Clustering of eastern North Pacific tropical cyclone tracks: ENSO and MJO effects. *Geochemistry, Geophysics and Geosystems*, **9**, Q06V05, doi: 10.1029/2007GC001861.

Camargo, S.J., M. Ting, and Y. Kushnir, 2012: Influence of local and remote SST on North Atlantic tropical cyclone potential intensity. *Climate Dynamics*

Camargo, S.J., M.K. Tippett, A.H. Sobel, G.A. Vecchi and M. Zhao, 2014: Testing the performance of tropical cyclone genesis indices in future climates using the HIRAM model. *J. Climate* (submitted)

Chavas, D.R. and K.A. Emanuel, 2010: A QuikSCAT climatology of tropical cyclone size. *Geophys. Res. Lett.*, **37**, L18816, doi:10.1029/2010GL044558

Chen, J.H., and S.J. Lin, 2011: The remarkable predictability of inter-annual variability of Atlantic hurricanes during the past decade. *Geophysical Research Letters*, **38** (L11804), doi:10.1029/2011GL047629.

Chen, J.-H., S.-J. Lin, 2013: Seasonal Predictions of Tropical Cyclones Using a 25-km-Resolution General Circulation Model. *J. Climate*, **26**, 380–398. doi: http://dx.doi.org/10.1175/JCLI-D-12-00061.1

Choi, K.-Y., G.A. Vecchi and A.T. Wittenberg, 2013: A mechanism for ENSO asymmetry. *J. Climate* doi: 10.1175/JCLI-D-13-00045.1.

Colbert, A.J., and B.J. Soden, 2012: *J. Climate*, **27**, 657.

Colbert, A.J., B.J. Soden, G.A. Vecchi and B.P. Kirtman, 2013: Impacts of Climate Change on North Atlantic Tropical Cyclone Tracks. J. Climate, 26, doi: 10.1175/JCLI-D-12-00342.1.

Dee, D.P., and coauthors, 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quart. J. Roy. Meteorol. Soc.*, **137**(656), 553-597.

Delworth, T.L., and coauthors, 2006: GFDL's CM2 Global Coupled Climate Models. Part I: Formulation and Simulation Characteristics. Journal of Climate, 19(5), DOI:10.1175/JCLI3629.1.

Delworth, T.L. and coauthors, 2012: Simulated climate and climate change in the GFDL CM2.5 high-resolution coupled climate model. J. Climate doi:10.1175/JCLI-D-11-00316.1

Doi, T., G.A. Vecchi, A.J. Rosati and T.L. Delworth, 2012: Tropical Atlantic biases in the mean state, seasonal cycle, and interannual variations for a coarse and high resolution coupled climate model. J. Climate, doi:10.1175/JCLI-D-11-00360.1

Elsner, J.B., B. H. Bossak, and X. F. Niu, 2001: Secular changes to the ENSO-U.S. hurricane relationship. Geophys. Res. Lett., 28, 4123–4126.

Emanuel, K. A., 1995: Sensitivity of tropical cyclones to surface exchange coefficients and a revised steady-state model incorporating eye dynamics, *J. Atmos. Sci.*, **52**, 3969–3976.

Emanuel, K. A., and D. S. Nolan, 2004: Tropical cyclones and the global climate system, paper presented at 26th Conference on Hurricanes and Tropical Meteorology, Am. Meteorol. Soc., Miami, Fla.

Emanuel, K.A., R. Sundararajan, and J. Williams, 2008: Hurricanes and global warming—Results from downscaling IPCC AR4 sim- ulations. *Bull. Amer. Meteor. Soc.*, **89**, 347–367.

Emanuel, K., Solomon, S., Folini, D., Davis, S. & Cagnazzo, C., 2013: Influence of tropical tropopause layer cooling on Atlantic hurricane activity. *J. Clim.*, **26**, 2288–2301.

Frank, W. M., and E. A. Ritchie, 2001: Effects of vertical wind shear on the intensity and structure of numerically simulated hurricanes, *Mon. Weather Rev.*, **129**, 2249–2269.

Gnanadesikan, A. et al., 2006: GFDL's CM2 global coupled climate models - Part 2: The baseline ocean simulation, *J. Climate*, **19**, 675-697.

Goddard, L., and Coauthors, 2013: A verification framework for interannual-to-decadal predictions experiments. *Climate Dyn.,* **40**, 245–272

Gray, William M., 1984: Atlantic Seasonal Hurricane Frequency. Part I: El Niño and 30 mb Quasi-Biennial Oscillation Influences. *Mon. Wea. Rev.*, **112**, 1649–1668.

Jagger, T. H., and J. B. Elsner, 2010: A consensus model for seasonal hurricane prediction. *J. Climate*, **23**, 6090–6099.

Jia, L. and Coauthors, 2014: Improved Seasonal Prediction of Temperature and Precipitation over Land in a High-resolution GFDL Climate Model, *J. Climate*, submitted.

Kam, J., J. Sheffield, X. Yuan, E.F. Wood, 2013: *J. Climate,* **26**, 3067.

Kim, H.-S., G.A. Vecchi, T.R. Knutson, W.G. Anderson, T.L. Delworth, A. Rosati, F. Zeng, M. Zhao, 2014: Tropical Cyclone Simulation and Response to CO2 Doubling in the GFDL CM2.5 High-Resolution Coupled Climate Model. *J. Climate* (submitted).

Klotzbach, P.J., and W.M. Gray, 2009: Twenty-five years of Atlantic basin seasonal hurricane forecasts, *Geophysical Research Letters*, **36** (L09711), doi:10.1029/2009GL037580.

Knapp, K.R., M.C. Kruk, D.H. Levinson, H.J. Diamond, and C.J. Neuman, 2010: The International Best Track Archive for Climate Stewardship (IBTrACS), *Bulletin of the American Meteorological Society*, **91**, 363-376.

Knutson, T.R., and coauthors, 2013: Dynamical Downscaling Projections of Late 21st Century Atlantic Hurricane Activity CMIP3 and CMIP5 Model-based Scenarios. *J. Climate*, doi:10.1175/JCLI-D-12-00539.1

Knutson, T.R., J.J. Sirutis, S.T. Garner, G.A. Vecchi and I.M. Held, 2008: Simulated reduction in Atlantic hurricane frequency under twenty-first-century warming conditions, *Nature Geoscience*, doi:10.1038/ngeo202

Kosaka, Y., S.-P. Xie, N.-C. Lau and G.A. Vecchi, 2013: An air-sea coupled mode over the Indo-Northwestern Pacific warm pool in boreal summer. *Proc. Nat. Acad. Sciences.* doi: 10.1073/pnas.1215582110

Kossin, J. P., and D. J. Vimont, 2007: A more general framework for understanding Atlantic hurricane variability and trends. Bull. Amer. Meteor. Soc., 88, 1767-1781.

Kossin, J. P., Camargo, S. J. and Sitkowski, M., 2010: Climate modulation of North Atlantic hurricane tracks. *J. Clim.,* **23**, 3057-3076.

Landsea, C.W., G.A. Vecchi, L. Bengtsson, and T.R. Knutson, 2009: Impact of Duration Thresholds on Atlantic Tropical Cyclone Counts. *J. Climate* doi: 10.1175/2009JCLI3034.1

LaRow, T. E., 2013: The Impact of SST Bias Correction on North Atlantic Hurricane Retrospective Forecasts. *Mon Wea Rev.*, **141**, 490-498.

LaRow, T. E., L. Stefanova, D. W. Shin and S. Cocke, 2010: Seasonal Atlantic Tropical Cyclone Hindcasting/Forecasting using Two Sea Surface Temperature Datasets. *Geophys. Res. Lett.*, **37**, L02804, doi:10.1029/2009GL041459.

Lengaigne, M. and G.A. Vecchi (2009). Contrasting the termination of moderate and extreme El Niño events in Coupled General Circulation Models. *Climate Dynamics*, doi:10.1007/s00382-009-0562-3

Lin, N., J.A. Smith, G. Villarini, T. P. Marchok, M.L. Baeck, 2010: Modeling Extreme Rainfall, Winds, and Surge from Hurricane Isabel (2003). *Wea. Forecasting*, **25**, 1342–1361. doi: http://dx.doi.org/10.1175/2010WAF2222349.1

Magnusson, L., M. Alonso-Balmaseda, S. Corti, F. Molteni, T. Stockdale, 2013: Evaluation of forecast strategies for seasonal and decadal forecasts in presence of systematic model errors. *Clim. Dyn.*, **41**(9-10), 2393-2409.

Msadek, R., K.W. Dixon, T.L. Delworth, and W.J. Hurlin, 2010: Assessing the predictability of the Atlantic meridional overturning circulation and associated fingerprints. *Geophys. Res. Lett.*, **37**, L19608, DOI:10.1029/2010GL044517.

Msadek, R., W.E. Johns, S.G. Yeager, G. Danabasoglu, T.L. Delworth, and A. Rosati, 2013: The Atlantic Meridional Heat transport at 26.5° N and its relationship with the MOC in the RAPID array and the GFDL and NCAR coupled models. *J. Clim.*, **26**(12), DOI:10.1175/JCLI-D-12-00081.1.

Msadek, R., T. Delworth; A. Rosati; W. Anderson; G.A. Vecchi; Y.-S. Chang; K. Dixon; R. Gudgel; B. Stern; A. Wittenberg; X. Yang; F. Zeng; R. Zhang; S. Zhang, 2014.a: Predicting a decadal shift in North Atlantic climate variability using the GFDL forecast system. *J. Climate (submitted)*.

Msadek R., G. A. Vecchi, and T. R. Knutson, 2014.b: North Atlantic Hurricane Activity: Past, Present and Future. Chapter for the NTU Climate Change Conference Proceedings Book, (in press).

Murakami, H. and B. Wang, 2010: Future change in North Atlantic tropical cyclone tracks: Projection by a 20-km-mesh global atmospheric model. *J. Climate*, **23**, 2699-2721.

Murakami, H., B. Wang and A. Kitoh, 2011: Future change in western North Pacific typhoons. Projections by a 20-km-mesh global atmospheric model. *J. Climate.*

Murakami, H., and Coauthors, 2012: Future Changes in Tropical Cyclone Activity Projected by the New High-Resolution MRI-AGCM. *J. Climate*, **25**, 3237-3260.

Murakami, H., B. Wang, T. Li, and A. Kitoh, 2013: Projected increase in tropical cyclones near Hawaii. *Nature Climate Change.*

Murakami, H., P.-C. Hsu, O. Arakawa, T. Li, 2014: Influences of model biases on projected future changes in tropical cyclone occurrence. *J. Climate* (submitted)

Peduzzi, P., B. Chatenoux, H. Dao, A. De Bono, C. Herold, J. Kossin, F. Mouton, and O. Nordbeck, 2012: Global trends in tropical cyclone risk, *Nature Climate Change*, **2**, 289-294.

Pielke, R. A. Jr and coauthors, 2008: Normalized hurricane damages in the United States: 1900–2005 *Nat. Hazard. Rev.*, 9, 29–42.

Putman, W. M., and S.-J. Lin, 2007: Finite-volume transport on various cubed-sphere grids. *J. Comput. Phys.*, **227**, 55–78.

Rienecker, M.M., M.J. Suarez, R. Gelaro, R. Todling, J. Bacmeister, E. Liu, M.G. Bosilovich, S.D. Schubert, L. Takacs, G.-K. Kim, S. Bloom, J. Chen, D. Collins, A. Conaty, A. da Silva, W. Gu, J. Joiner, R.D. Koster, R. Lucchesi, A. Molod, T. Owens, S. Pawson, P. Pegion, C.R. Redder, R. Reichle, F.R. Robertson, A.G. Ruddick, M. Sienkiewicz, and J. Woollen, 2011: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications, *Journal of Climate*, **24**(14), 3624-3648.

Scocimarro, E., S. Gualdi, G. Villarini, G.A. Vecchi, M. Zhao, K. Walsh, A. Navarra, 2014: Intense precipitation events associated with landfalling tropical cyclones in a warmer climate. *J. Climate* (submitted).

Smith, D.M., R. Eade, N. J. Dunstone, D. Fereday, J. M. Murphy, H. Pohlmann, and A. A. Scaife, 2010: Skillful multi-year predictions of Atlantic hurricane frequency. *Nat. Geosci.*, **3**, 846–849.

Song, Q., G.A. Vecchi and A. Rosati, 2008: Predictability of Indian Ocean Sea Surface Temperature Anomalies in the GFDL Coupled Model. *Geophys. Res. Lett.* **5**, L02701, doi:10.1029/2007GL031966.

Swanson, K. L., 2008: Nonlocality of Atlantic tropical cyclone in- tensities. *Geochem. Geophys. Geosyst.*, **9**, Q04V01, doi:10.1029/ 2007GC001844.

Tippett, M. K., S. J. Camargo, and A. H. Sobel, 2011: A Poisson regression index for tropical cyclone genesis and the role of large-scale vorticity in genesis. *J. Climate*, **24**, 2335–2357.

Vecchi, G.A., 2006: The termination of the 1997-98 El Niño. Part II: Mechanisms of Atmospheric Change. *J. Climate*, **19**(12), 2647-2664.

Vecchi, G.A., and D.E. Harrison, 2006: The termination of the 1997-98 El Niño. Part I: Mechanisms of Oceanic Change. *J. Climate*, **19**(12), 2633-2646.

Vecchi, G.A., A.T. Wittenberg and A. Rosati, 2006: Reassessing the role of stochastic forcing in the 1997-8 El Niño. *Geophys. Res. Lett.* **33**, L01706, doi:10.1029/2005GL024738.

Vecchi, G.A., and B.J. Soden, 2007: Effect of remote sea surface temperature change on tropical cyclone potential intensity, *Nature*, **450**, 1066-1070 doi:10.1038/nature06423.

Vecchi, G.A., K.L. Swanson, and B.J. Soden, 2008: Whither Hurricane Activity? *Science* **322**(5902), 687. DOI: 10.1126/science.1164396

Vecchi, G.A., and T.R. Knutson, 2011: Estimating annual numbers of Atlantic hurricanes missing from the HURDAT database (1878-1965) using ship track density. *J. Climate.* doi: 10.1175/2010JCLI3810.1.

Vecchi, G.A., M. Zhao, H. Wang, G. Villarini, A. Rosati, A. Kumar, I. M. Held, and R. Gudgel, 2011: Statistical-dynamical predictions of seasonal North Atlantic hurricane activity, *Monthly Weather Review*, **139**(4), 1070-1082.

Vecchi, G.A., R. Msadek, W. Anderson, Y.-S. Chang, T. Delworth, K. Dixon, R. Gudgel, A. Rosati, W. Stern, G. Villarini, A. Wittenberg, X. Yang, F. Zeng, R. Zhang, and S. Zhang, Multi-year predictions of North Atlantic hurricane frequency: Promise and limitations, *Journal of Climate*, **26**(15), 5337-5357, 2013.a.

Vecchi, G.A., S. Fueglistaler, I.M. Held, T.R. Knutson, and M. Zhao, 2013.b: Impacts of Atmospheric Temperature Changes on Tropical Cyclone Activity. *J. Climate*. doi: 10.1175/JCLI-D-12-00503.1

Vecchi, G.A. and G. Villarini, 2014: Enhancing Seasonal Hurricane Predictions. *Science*, doi: 10.1126/science.1247759

Villarini, G., G.A. Vecchi, and J.A. Smith, 2010: Modeling of the dependence of tropical storm counts in the North Atlantic Basin on climate indices, *Monthly Weather Review*, **138**(7), 2681-2705.

Villarini, G., G.A. Vecchi, T.R. Knutson and J.A. Smith, 2011.a: Is the Recorded Increase in Short Duration North Atlantic Tropical Storms Spurious? *J. Geophys. Res.* doi:10.1029/2010JD015493.

Villarini, G., J.A. Smith, M.L. Baeck, T. Marchok, and G.A. Vecchi, 2011.b: Analysis of Rainfall Distribution for U.S. Landfalling Tropical Cyclones: Frances, Ivan and Jeanne (2004). *J. Geophys. Res.* doi:10.1029/2011JD016175

Villarini, G., G.A. Vecchi, and J.A. Smith, 2012: U.S. landfalling and North Atlantic hurricanes: Statistical modeling of their frequencies and ratios, *Monthly Weather Review*, **140**(1), 44-65.

Villarini, G., and G.A. Vecchi, 2013: Multi-Season Lead Forecast of the North Atlantic Power Dissipation Index (PDI) and Accumulated Cyclone Energy (ACE). *J. Climate*, doi:10.1175/JCLI-D-12-00448.

Villarini, G., R. Goska, J.A. Smith and G.A. Vecchi, 2014.a: North Atlantic Tropical Cyclones and U.S. Flooding. *Bull Amer. Meteorol. Soc.* (in press).

Villarini, G., D.A. Lavers, E. Scocimarro, M. Zhao, M.F. Wehner, G.A. Vecchi and T.R. Knutson, 2014.b: Sensitivity of Tropical Cyclone Rainfall to Idealized Global Scale Forcing. *J. Climate* (submitted).

Vimont, D. J., and J. P. Kossin, 2007: The Atlantic Meridional Mode and hurricane activity. *Geophys. Res. Lett.*, **34**, L07709, doi:10.1029/2007GL029683.

Vitart, F., J. L. Anderson, W. F. Stern, 1997: Simulation of Interannual Variability of Tropical Storm Frequency in an Ensemble of GCM Integrations. *J. Climate*, **10**, 745–760.

Vitart, F., and T.N. Stockdale, 2001: Seasonal forecasting of tropical storms using coupled GCM integrations. *Mon. Wea. Rev*, **129**(10), 2521-2527.

Vitart, F., 2006: Seasonal forecasting of tropical storm frequency using a multi-model ensemble. *Quart. J. Roy. Meteor. Soc.*, **132**, 647–666.

Vitart, F., M. Huddleston, M. Deque, D. Peake, T. Palmer, T. Stockdale, M. Davey, S. Ineson and A. Weisheimer, 2007: Dynamically-based seasonal forecast of Atlantic tropical storm activity issued in June by EUROSIP. *Geophys. Res. Lett.*, **34**, L16815, doi:10.1029/2007GL030740.

Wang, H., J.K.E. Schemm, A. Kumar, W. Wang, L. Long, M. Chelliah, G.D. Bell, and P. Peng, 2009: A statistical forecast model for Atlantic seasonal hurricane activity

based on the NCEP dynamical seasonal forecast, *Journal of Climate*, **22**, 4481-4500.

Wang, H. and coauthors, 2014: How Well Do Global Climate Models Simulate the Variability of Atlantic Tropical Cyclones Associated with ENSO? *J. Climate* (submitted).

Wittenberg, A.T., A. Rosati, N.-C. Lau, and J.J Ploshay, 2006: GFDL's CM2 Global Coupled Climate Models. Part III: Tropical Pacific Climate and ENSO. *J. Climate*, **19**(5), DOI:10.1175/JCLI3631.1.

Wittenberg, Andrew T., 2009: Are historical records sufficient to constrain ENSO simulations? *Geophysical Research Letters*, **36**, L12702, DOI:10.1029/2009GL038710.

Wittenberg, A.T., A. Rosati, T.L. Delworth, G.A. Vecchi and F. Zeng, 2014.a: Interdecadal ENSO modulation: Is it predictable? *J. Climate* (in press). DOI:10.1175/JCLI-D-13-00577.1.

Wittenberg, A.T., and coauthors, 2014.b: Tropical Pacific Climate and Variability in the GFDL Global Coupled GCMs: Sensitivities to Resolution, Formulation, and Bias Correction. *J. Climate (to be submitted).*

Yang., X. and co-authors, 2013: A predictable AMO-like pattern in GFDL's fully-coupled ensemble initialization and decadal forecasting system. *J. Climate*, doi:10.1175/JCLI-D-12-00231.

Zhang W., H. Graf, Y. Leung, M. Herzog, 2012: Different El Niño Types and Tropical Cyclone Landfall in East Asia. *J. Climate*, **25**, 6510-6523.

Zhang, W., Y. Leung, and J.C.L. Chan, 2013.a: The analysis of tropical cyclone tracks in the western North Pacific through data mining, Part I: tropical cyclone recurvature. *J. Applied Meteorology and Climatology*, **52**, 1394-1416.

Zhang, W., Y. Leung, and J.C.L. Chan, 2013.b: The analysis of tropical cyclone tracks in the western North Pacific through data mining, Part II tropical cyclone landfall. *J. Applied Meteorology and Climatology*, **52**, 1417-1432.

Zhang, W., Y. Leung, and Y. Wang, 2013.c: Cluster analysis of Post-landfall Tracks of Landfalling Tropical Cyclones over China. *Climate Dynamics*, 1237-1255.

Zhao, M., I.M. Held, S.-J. Lin, and G.A. Vecchi, 2009: Simulations of global hurricane climatology, interannual variability, and response to global warming using a 50km resolution GCM. *J. Climate*, **22**(24), 6653-6678, doi:10.1175/2009JCLI3049.1

Zhao, M., I.M. Held, and G.A. Vecchi, 2010: Retrospective forecasts of the hurricane season using a global atmospheric model assuming persistence of SST anomalies. *Mon.Wea. Rev*., **138**, 3858–3868.

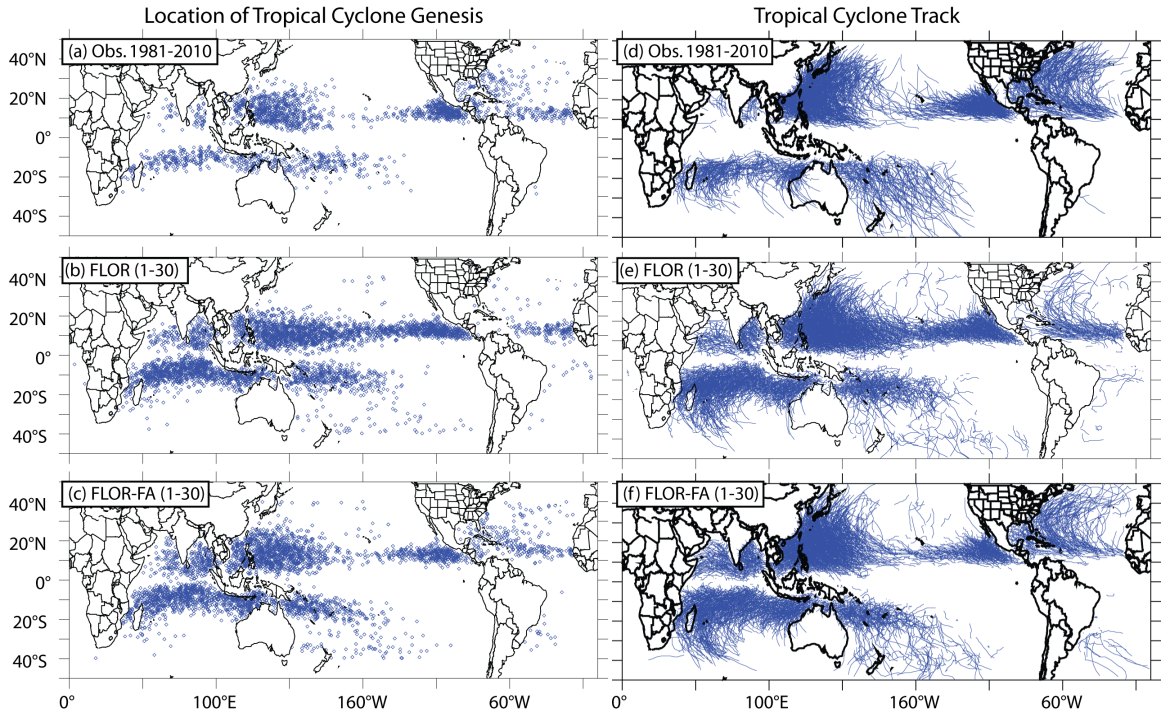

Figure 1: Location of tropical cyclone genesis (left column) and tropical cyclone tracks (right column). The top panels are based on the IBTRACS data (Knapp et al. 2010), focusing on TCs that last a minimum of two days. The middle panels are based on the first 30 years of a control climate simulation with 1990 radiative forcing values from FLOR. The bottom panels are based on the first 30 years of a control climate simulation with 1990 radiative forcing values from FLOR-FA.

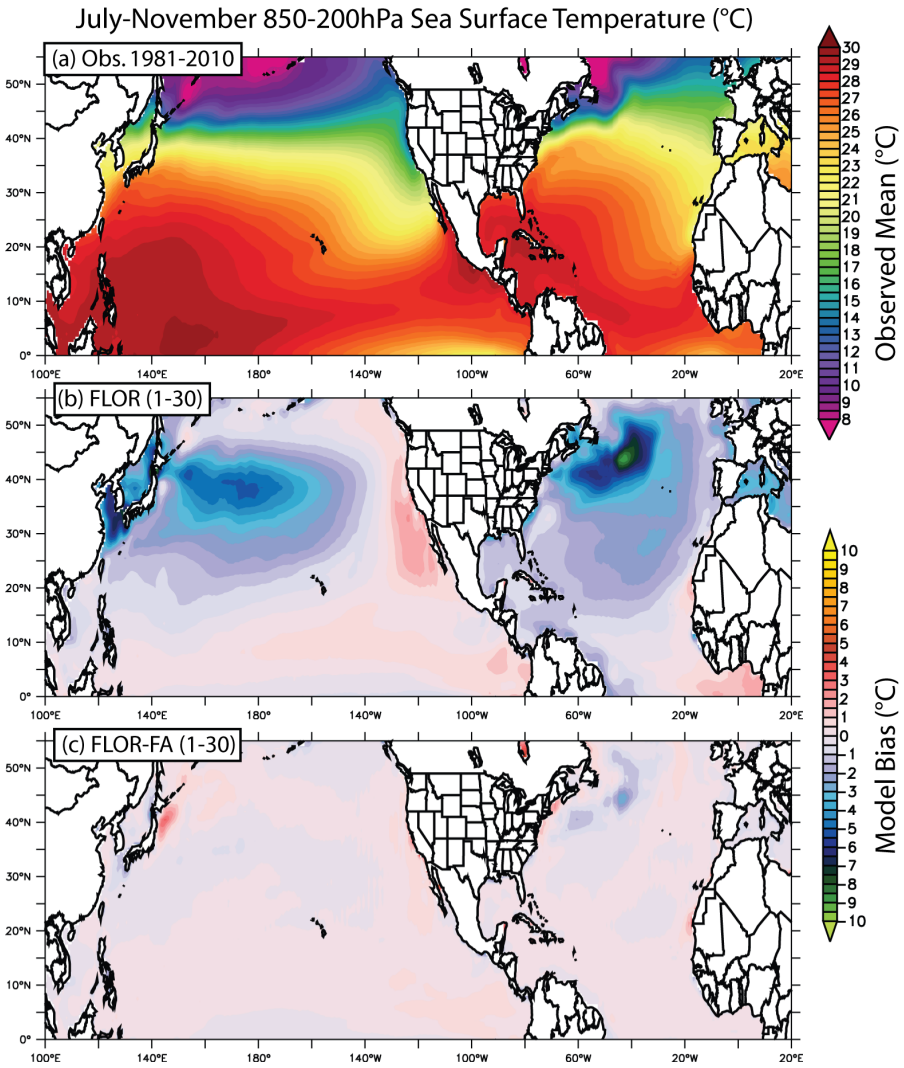

Figure 2: Bias and impact of flux adjustment on SST. Top panels show the climatological July-November values computed over 1981-2010 from the HadISST.v1 SST product (Rayner et al. 2004); results are very similar when comparing against SSTs from either the ERA-I (Dee *et al.* 2011) or MERRA analyses (Rienecker et al. 2011). The middle and lower panels show the difference between the climatological July-November values averaged over the first 100 years of a the Present Day Control simulation with FLOR and FLOR-FA, respectively. Notice the reduction of bias arising from flux adjustment.

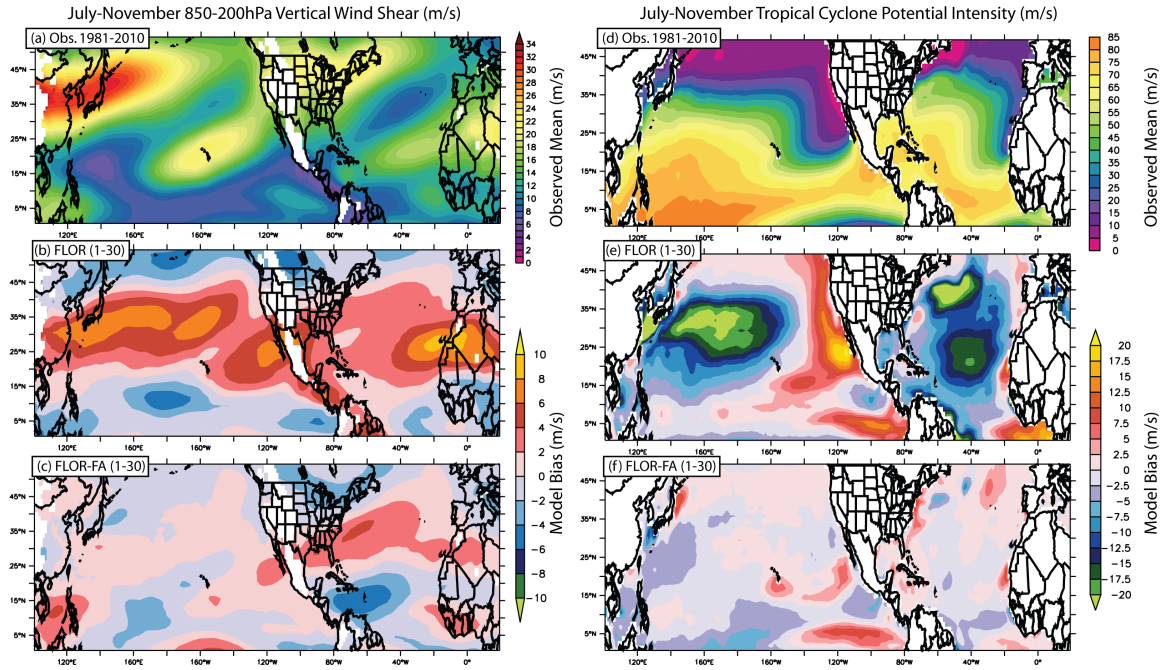

Figure 3: Bias and impact of flux adjustment on vertical wind shear (left column) and tropical cyclone potential intensity (right column). Top panels show the climatological July-November values computed over 1981-2010 from the MERRA analysis (Rienecker et al. 2011), results are very similar when compared to output from the ERA-Interim reanalysis (Dee *et al.* 2011; not shown). The middle and lower panels show the difference between the MERRA climatological July-November values averaged over the first 100 years of a the Present Day Control simulation with FLOR and FLOR-FA, respectively. Notice the reduction of bias arising from flux adjustment.

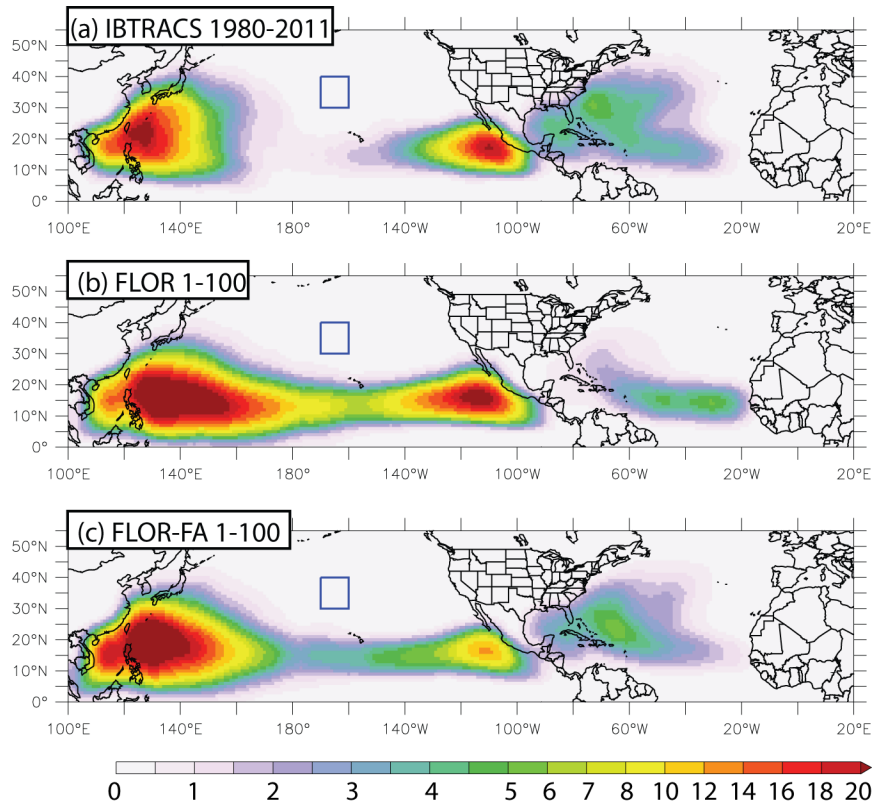

Figure 4: Long-term average 10°×10° TS density. Panel (a) shows 1980-2011 average from IBTRACS (Knapp et al. 2010), while panels (b) and (c) show the average over the first 100 years of the Present Day Control simulation with FLOR and FLOR-FA, respectively. Blue box in the northern central Pacific indicates the 10°×10° scale, for reference.

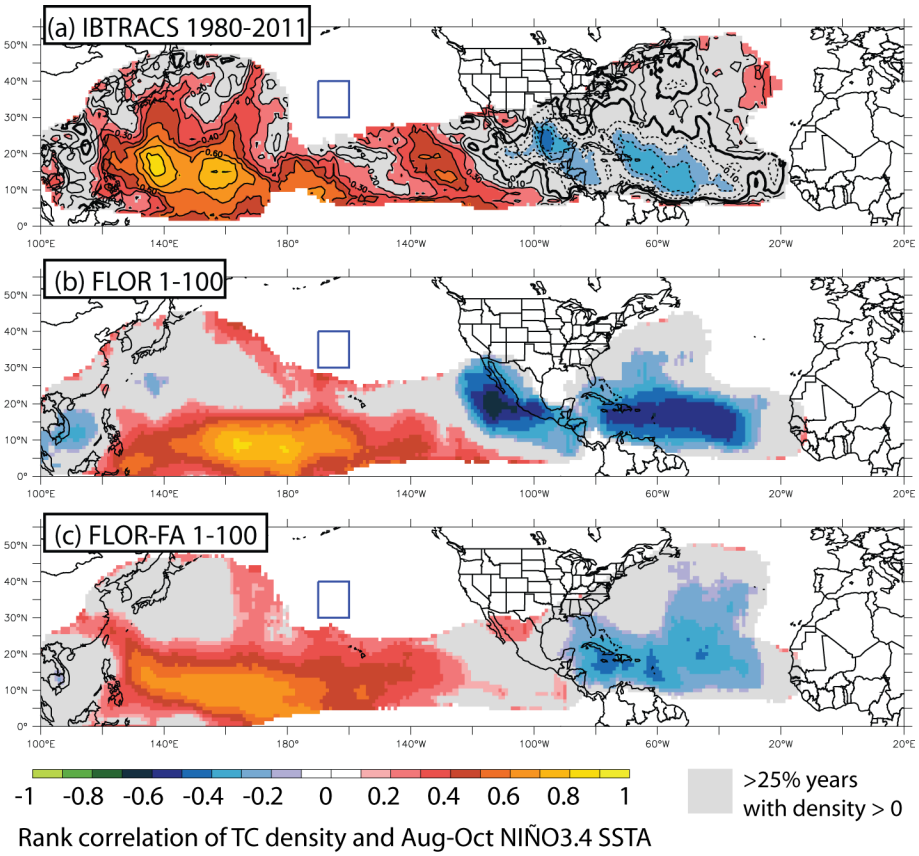

Figure 5: Rank correlation of 10°×10° TS density and August-October NIÑO3.4 SST Anomalies. Panel (a) shows 1980-2011 rank correlation of TC density from IBTRACS (Knapp et al. 2010) against NIÑO3.4 SSTA from HadISST (Rayner et al. 2004), and panels (b) and (c) show the rank correlations over the first 100 years of the Present Day Control simulation with FLOR and FLOR-FA, respectively. In panel (a) the rank correlation is masked at the p=0.2 level, with non-significant values shown by unshaded contours. In panels (b) and (c) the rank correlation is masked at the p=0.1 level. Gray shading in all panels indicates the regions of the northern Pacific and Atlantic in which TC density for each dataset is non-zero for 25% of the years. Blue box in the northern central Pacific indicates the 10°×10° scale, for reference.

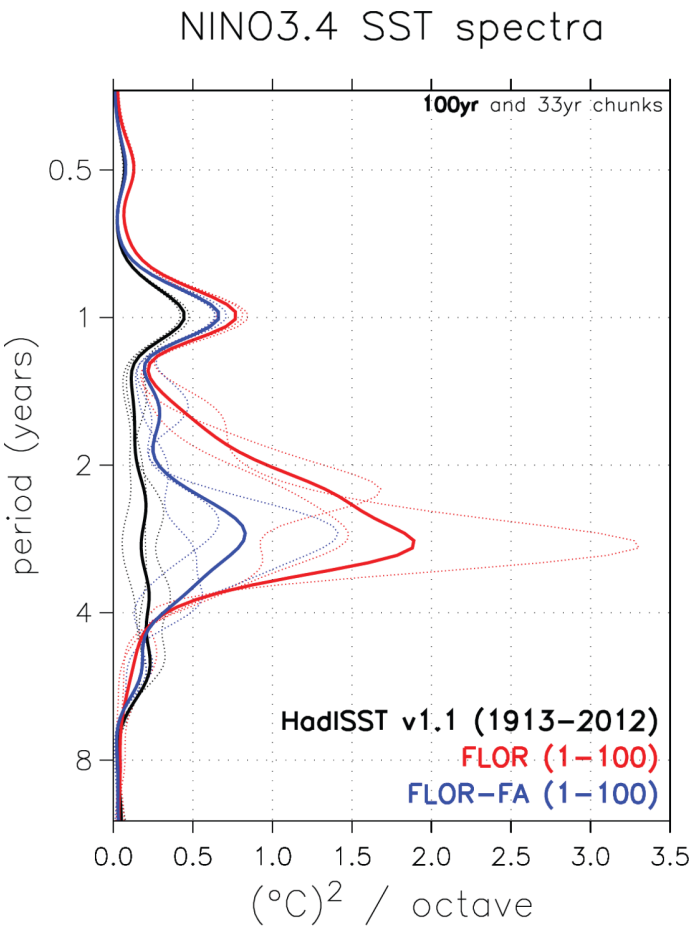

Figure 6: Power spectrum of NIÑO3.4 SST (170°W-120°W, 5°S-5°N average) based on observational estimates, FLOR and FLOR-FA. Black line shows the power spectrum computed from monthly NIÑO3.4 SST using the UK Met Office Hadley Center Interpolated SST Product (HadISSTv1, Rayner et al. 2004). In the red (blue) line shows the power spectrum computed from monthly NIÑO3.4 SST from FLOR (FLOR-FA). Thick lines show the values over a 100-year segment, thin lines show the values over non-overlapping 33-year segments to highlight variability.

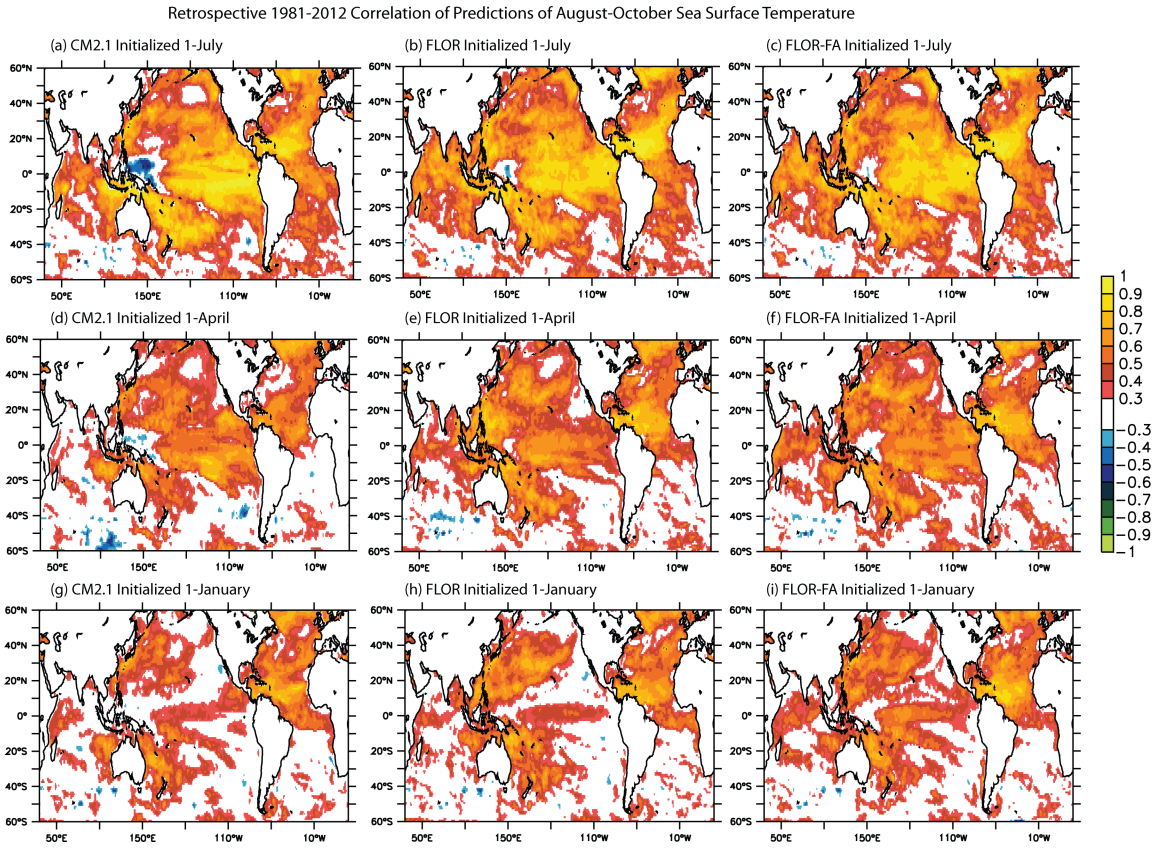

Figure 7: Retrospective forecast skill of 1981-2012 August-October SST. Shading indicates the retrospective correlation of predicted vs. observed SST (HadISST, Rayner et al. 2004). Top panels focus on the 1-July initialized forecasts, middle panels focus on the 1-April initialized forecasts, and bottom panels focus on the 1-January initialized forecasts. Left panels show the results for CM2.1, middle panels for FLOR, right panels for the flux adjusted version of FLOR.

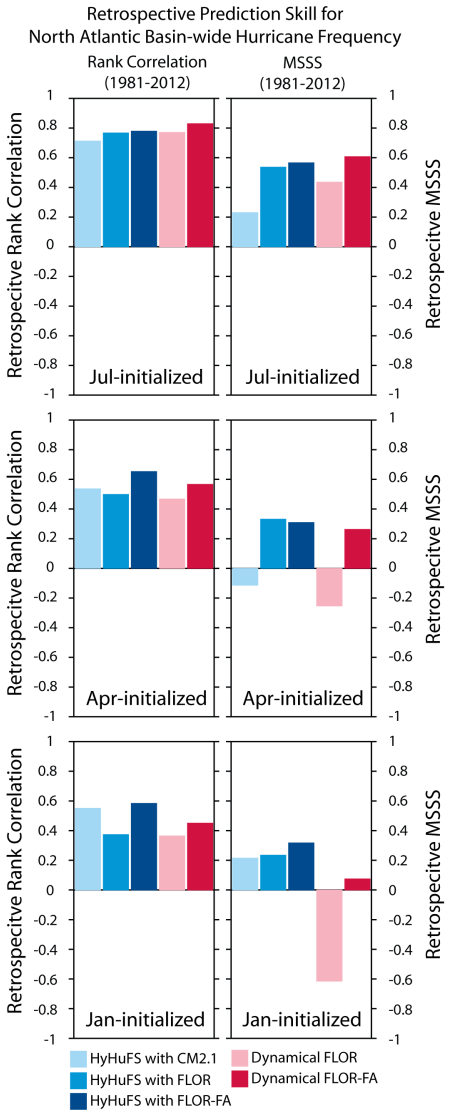

Figure 8: 1981-2012 retrospective forecast skill for North Atlantic hurricane frequency using hybrid statistical dynamical (blue bars) and dynamical (red bars) approaches. Top, middle and bottom panels show the values for 1-July, 1-April and 1-January initialized forecasts. Left panels show the retrospective rank correlation between the predictions and observations (IBTRACS, Knapp et al. 2010), and right panels show the retrospective mean skill score square against observations (IBTRACS, Knapp et al. 2010).

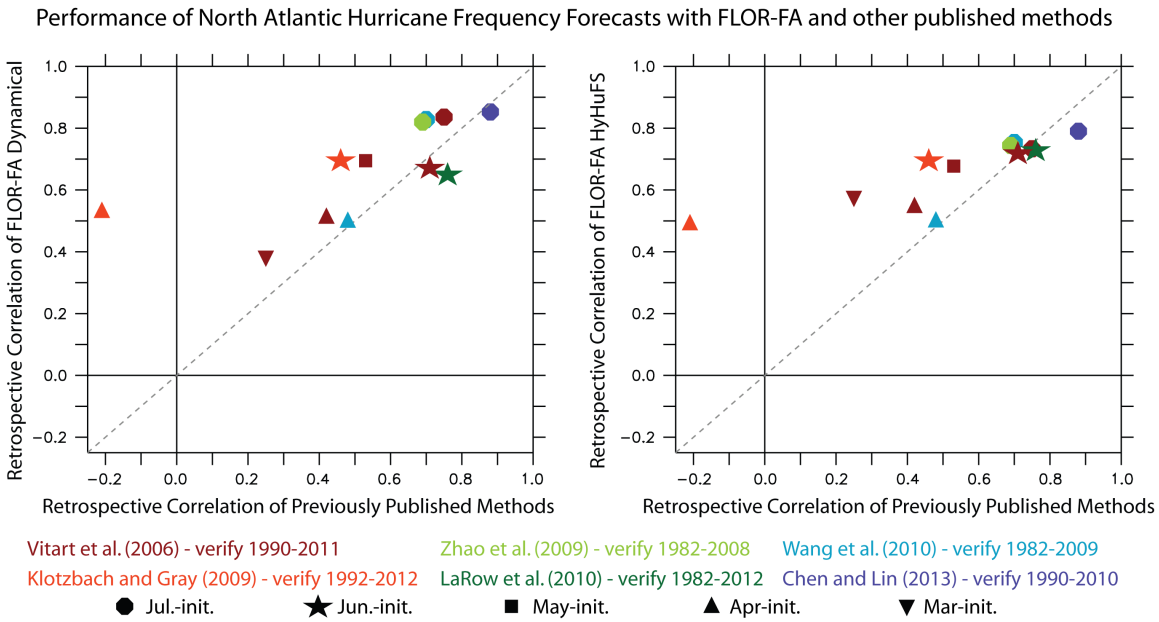

Figure 9: Comparison of retrospective forecast skill for North Atlantic hurricane frequency from the dynamical (left panel) and HyHuFS (right panel) seasonal predictions with FLOR-FA with those of other methodologies in the published literature. In this figure, in contrast to Fig. 8, we use the Pearson correlation against observed hurricane frequency as our measure of skill, since it was that used in these other published studies. The correlations with FLOR-FA are computed over the same periods as each published study, with the colors of the symbols indicating the method to which FLOR-FA is being compared. The various symbols indicate the initialization month for the predictions. The 1:1 line is indicated in dashed gray.

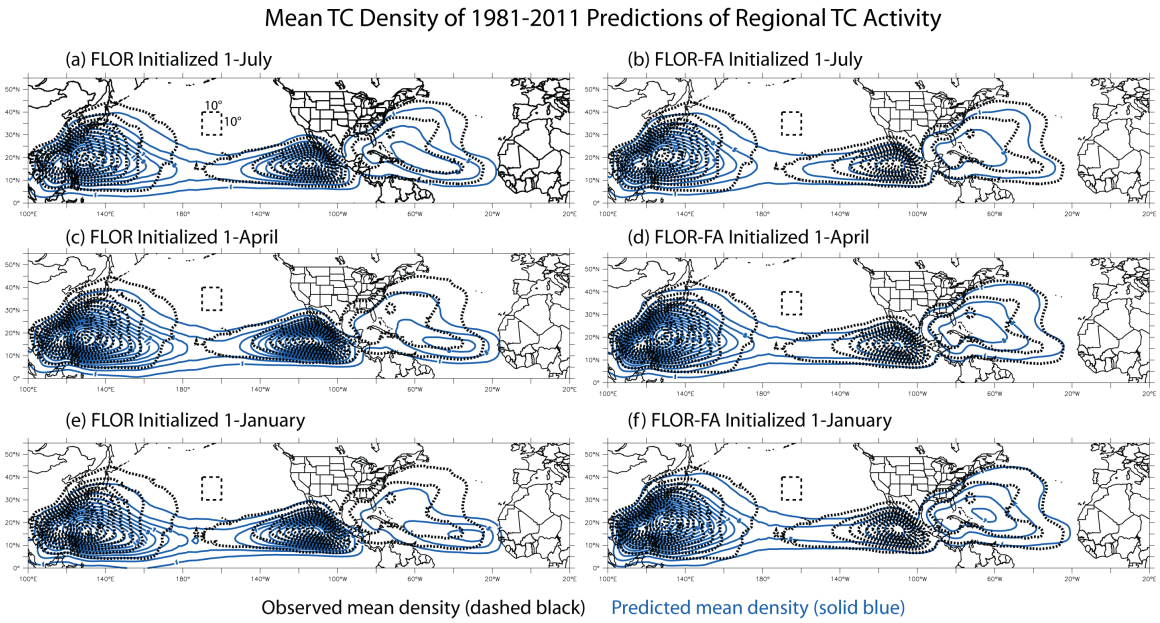

Figure 10: Mean observed (black dotted; Knapp et al. 2010) and predicted (blue) 10°×10° TC density based over 1981-2011. Top panels focus on the 1-July initialized forecasts, middle panels focus on the 1-April initialized forecasts, and bottom panels focus on 1-January initialized forecasts. Left panels show the results for FLOR, right panels for the flux adjusted version of FLOR. For each forecast, density computed for the post-initialization months of the calendar year for both observations and the forecasts (*e.g.,* for 1-July they are based on July-December). Dashed box in the northern central Pacific indicates the 10°×10° scale, for reference.

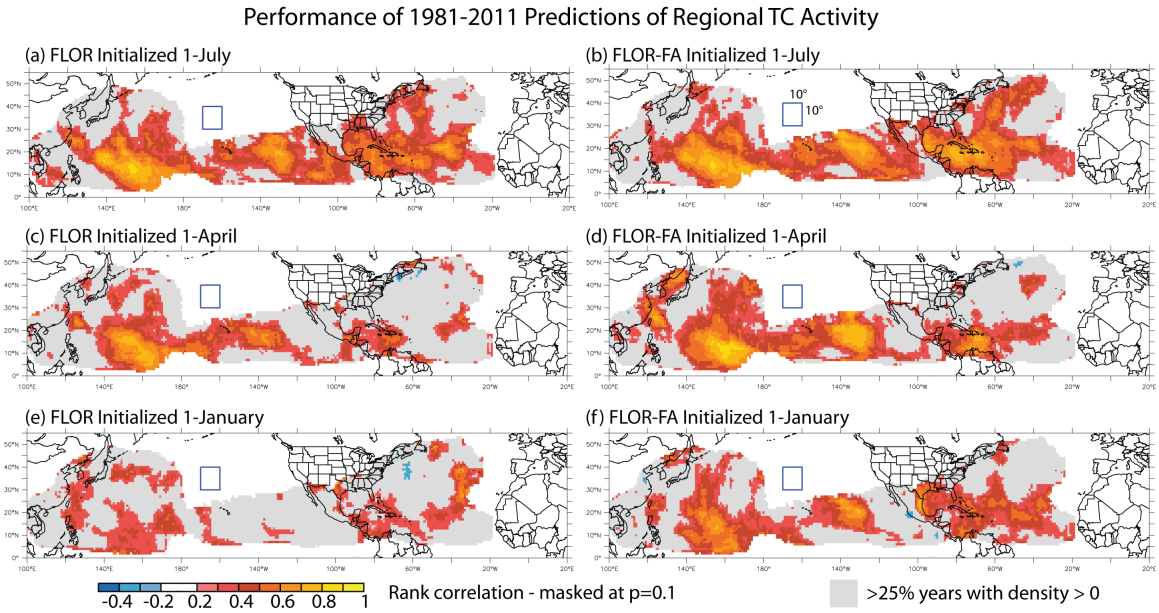

Figure 11: Retrospective forecast skill of 1981-2011 tropical cyclone density. Shading indicates the retrospective rank correlation of predicted vs. observed (IBTRACS, Knapp et al. 2010) 10°×10° TC density, masked at a two-sided *p=0.1* level. Top panels focus on the 1-July initialized forecasts, middle panels focus on the 1-April initialized forecasts, and bottom panels focus on 1-January initialized forecasts. Left panels show the results for FLOR, right panels for the flux adjusted version of FLOR. For each forecast, density computed for the post-initialization months of the year for both observations and the forecasts (*e.g.,* for 1-July they are based on July-December). Blue box in the northern central Pacific indicates the 10°x10° scale, for reference. Gray shading in all panels indicates the regions of the northern Pacific and Atlantic in which observed TC density is non-zero for at least 25% of the years.

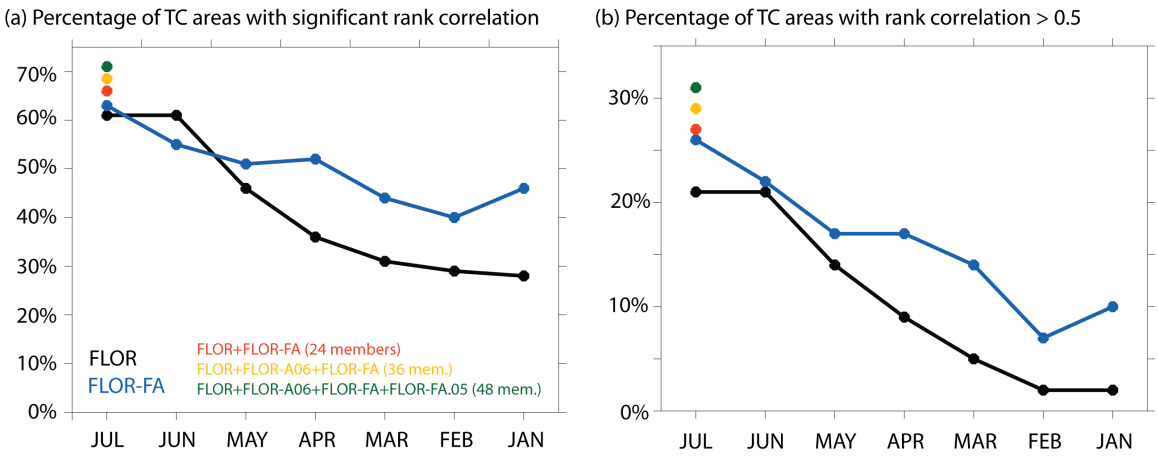

Figure 12: Percent area of North Pacific and Atlantic regions with non-zero observed 10°×10° TC density in one quarter of available years that show over 1981-2011, when compared with observations (IBTRACS, Knapp et al. 2010): (a) significant (at $p<0.1$) retrospective forecast rank correlation, and (b) retrospective rank correlation greater than 0.5. The different lines show the values using FLOR (black) and FLOR-FA (blue), the horizontal axis indicates the initialization dates (longest lead forecasts are to the right). For the July-initialized forecasts, the red, yellow and green symbols indicate the retrospective skill measures for the FLOR+FLOR-FA 24-member ensemble, FLOR+FLOR-FA+FLOR-A06 36-member ensemble, and the FLOR+FLOR-FA+FLOR-A06+FLOR-FA.05 48-member ensemble.

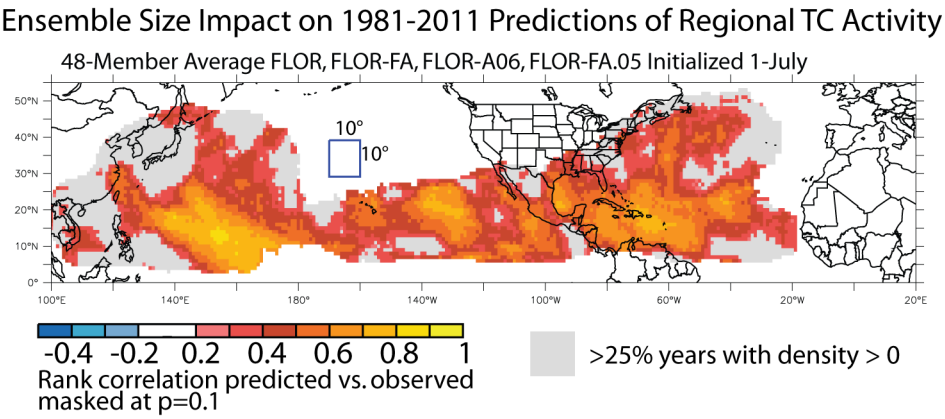

Figure 13: Retrospective forecast skill of 1981-2011 tropical cyclone density combining different versions of FLOR and FLOR-FA initialized 1-July to create a larger ensemble size. Shading indicates the retrospective rank correlation of predicted vs. observed (IBTRACS, Knapp et al. 2010) 10°×10° TC density, masked at a two-sided *p=0.1* level. Results are based on a 48-member ensemble created by combining the 12-member ensembles of predictions with FLOR, FLOR-FA, FLOR-A06 and FLOR-FA.05; see Section 2.a and 2.b for descriptions of the models). TS density computed over July-December. Blue box in the northern central Pacific indicates the 10°×10° scale, for reference. Gray shading indicates the regions of the northern Pacific and Atlantic in which observed TC density is non-zero for at least 25% of the years.